TEXT FLY WITHIN
THE BOOK ONLY

UNIVERSAL
LIBRARY

OU_160079

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 658/S74M Accession No. G. 10466

Author Spencer, Milton H., and Siegelman, Louis

Title Managerial economics... 1959.

This book should be returned on or before the date last marked below.

THE IRWIN SERIES IN ECONOMICS

CONSULTING EDITOR

LLOYD G. REYNOLDS

YALE UNIVERSITY

BOOKS IN THE IRWIN SERIES IN ECONOMICS

Anderson, Gitlow & Diamond (Editors) *General Economics: A Book of Readings*

Bushaw & Clower *Introduction to Mathematical Economics*

Cartter *Theory of Wages and Employment*

Davidson, Smith & Wiley *Economics: An Analytical Approach*

Due *Government Finance: An Economic Analysis*

Due *Intermediate Economic Analysis* Third Edition

Dunlop & Healy *Collective Bargaining: Principles and Cases* Revised Edition

Grampp & Weiler (Editors) *Economic Policy: Readings in Political Economy* Revised Edition

Grossman, Hansen, Hendriksen, McAllister, Okuda, & Wolman (Editors) *Readings in Current Economics*

Guthrie *Economics*

Halm *Economics of Money and Banking*

Harriss *The American Economy: Principles, Practices, and Policies* Revised Edition

Jome *Principles of Money and Banking*

Kindleberger *International Economics* Revised Edition

Lee *Economic Fluctuations*

Locklin *Economics of Transportation* Fourth Edition

Meyers *Economics of Labor Relations*

Scitovsky *Welfare and Competition: The Economics of a Fully Employed Economy*

Shaw *Money, Income, and Monetary Policy*

Snider *Introduction to International Economics* Revised Edition

Spencer & Siegelman *Managerial Economics: Decision Making and Forward Planning*

Wilcox *Public Policies Toward Business*

MANAGERIAL ECONOMICS

Decision Making and *Forward* Planning

MANAGERIAL ECONOMICS

DECISION MAKING
AND *FORWARD* PLANNING

BY

MILTON H. SPENCER, Ph.D.

ASSOCIATE PROFESSOR OF BUSINESS ADMINISTRATION
WAYNE STATE UNIVERSITY, DETROIT

AND

LOUIS SIEGELMAN, Ph.D.

THE FIRST NATIONAL BANK OF CHICAGO AND
LECTURER IN FINANCE, NORTHWESTERN
UNIVERSITY SCHOOL OF BUSINESS

1959

RICHARD D. IRWIN, INC.

HOMEWOOD, ILLINOIS

© 1959 BY RICHARD D. IRWIN, INC.

ALL RIGHTS RESERVED. THIS BOOK OR ANY PART
THEREOF MAY NOT BE REPRODUCED WITHOUT
THE WRITTEN PERMISSION OF THE PUBLISHER

First Printing, January, 1959

PRINTED IN THE UNITED STATES OF AMERICA

The Library of Congress has cataloged this book as follows:

Spencer, Milton H Managerial economics; decision making and forward planning, by Milton H. Spencer and Louis Siegelman. Homewood, Ill., R. D. Irwin, 1959. 454 p. illus. 24 cm. (The Irwin series in economics) 1. Industrial management. 2. Decision-making. 3. Business forecasting. 4. Economics, Mathematical. I. Siegelman, Louis, joint author. II. Title. HD31.S62 330.182 59-6857‡

To Our Parents
who made this book possible
and Our Wives
who made it necessary

PREFACE

This book is aimed primarily at three classes of readers: (1) upper division and graduate students in economics and business administration to whom an integrated training in these areas can be of value as professional researchers and potential executives; (2) academicians and business consultants who will find in these pages a good indication of the type of business research that can be and is being done in American industry today; and (3) executives who, as professional decision makers, usually have neither the time nor the training to perform the technical research needed to guide their actions, and hence may discover from reading this book the advantages to be gained by employing the services of trained business economists.

The book is econometrically oriented but nonmathematical in its presentation. Emphasis is placed on approaching the problems of management decision making and forward planning by formulating problems in a conceptually quantitative manner capable of numerical solution. Occasional calculations and graphic techniques are employed, but these can be readily understood by anyone possessing a knowledge of "advanced arithmetic." Thus, one of the unique features of the book is an extended treatment of graphic (multiple) correlation in Chapter 3 and a discussion of the role of correlation analysis in economic measurement and forecasting. The typical student in business or economics rarely goes beyond simple linear correlation in a standard statistics course, and it is surprising how much more he gains in interpreting the econometric studies discussed later in the book once he has at least a graphic understanding of correlation analysis.

In general, the book is sufficiently self-contained so that no particular advanced course or set of advanced courses is necessary as a prerequisite. Since the book will probably be used mainly at the senior or graduate level, students will already have a sufficient background in general economics or business administration to carry them safely through. Basically, the book attempts to integrate economic principles with various areas of business administration, and each student will find different fields to challenge him, depending on his own educational background. Thus, the student who knows some accounting should grasp more readily the discussion of profit measurement; the student who has had corporation finance may find parts of the discussion of capital management relatively easy; and the student who has had economic theory may find less difficulty with the treatment of production functions, and so on. In any event, our own experience has been this: by concentrating on the questions at the end of each chapter as a basis for class discussion (rather than lecturing directly from the book), the burden of reading is placed where it belongs—directly on the student—and it soon becomes apparent as to where the class as a whole seems to be weakest. The lectures can then be amplified accordingly.

To minimize the use of footnotes, and at the same time provide a guide for supplementary reading, a bibliographical note has been appended to each chapter. These notes are usually divided into two sections: one of a more technical and specialized nature that should be of use to teachers and advanced readers, and another which is more comprehensive and suitable for supplementary reading by the average student. Although these bibliographical notes are by no means exhaustive, we think that they are reasonably representative of the literature in the field.

We wish to take this opportunity to express an everlasting intellectual debt and our deepest appreciation to Professor Charles R. Whittlesey of the University of Pennsylvania, and Professor Arthur D. Gayer, late of Queens College. Their friendly encouragement and help, their stimulating instruction, and their counsel were essential factors in providing direction and channeling our interests along the route which led to the writing of this book. To Professor Simon Kuznets of The Johns Hopkins University, Professor Morris A. Copeland of Cornell University, and Dr. Karl Bopp, President of the Federal Reserve Bank of Philadelphia, we express our thanks for the inspiration we were privileged to draw from them during our formative years.

Our appreciation is also extended to various teachers, friends, and associates whose influence on our thinking is reflected in these pages, and to those whose assistance smoothed the way to bringing this book to fulfillment. We refer in particular to Dean Walter C. Folley and Professor James R. Taylor of Wayne State University; to Messrs. Herbert V. Prochnow and Joseph T. Keckeisen, Vice Presidents of The First National Bank of Chicago; to Professors Melvin de Chazeau and M. Slade Kendrick of Cornell University; to Mr. Werner Liebert, Vice President of Marketing, DWG Cigar Corporation; and to Mr. Ray Ayer, Manager of Market Analysis for Dodge Division, Chrysler Corporation. Special mention is also made of Mr. Edwin Kalmore of the public accounting firm of S. D. Leidesdorf, Mr. Samuel Kam of Stern Brothers Department Store, and Mr. Norman Weinstein of Weinstein Liquors.

Professor Robert Clower of Northwestern University read the manuscript and offered many helpful suggestions. Messrs. Thomas H. Beacom, Vice President, and Tom M. Plank, Assistant Cashier, The First National Bank of Chicago, read portions of the manuscript in its final stages. To these people we express our thanks, though we must, of course, accept all criticism for any incorrect statements which remain.

The Econometric Institute, Inc. very kindly gave us permission to reproduce the final results of several of their studies. The Iowa State College Press, and the McGraw-Hill Publishing Company, Inc., were also very kind to permit us to reproduce some of their copyrighted materials. The cooperation of these and other copyright owners, to whom credit is given in the text, is here gratefully acknowledged.

To those frequently unsung heroes and heroines of college texts—the librarians, secretaries, and student assistants—we owe our greatest direct debt. Of particular importance in this respect are Miss Marion E. Wells, Librarian, and Miss Martha A. Whaley, Assistant Librarian, The First National Bank of Chicago; and Mrs. Virginia Bandoni, Mr. Tom Boufford, and Mr. Si Mikaelian whose efforts were so greatly appreciated during those very hectic days prior to final examinations, summer vacation, and the promised delivery date to the publisher. Substantial credit also goes to Mrs. Grace Mattheiss, Mrs. Bea Merians, and Mrs. Barbara Schore for their cheerful cooperation during the manuscript's early stages. To all of these people we offer our sincere thanks, and we hope that the day will soon come when society provides more just rewards for the services of perfect librarians and secretaries.

MHS
LS

TABLE OF CONTENTS

PART I. UNCERTAINTY AND PREDICTION

CHAPTER PAGE

1 THE UNCERTAINTY FRAMEWORK OF MANAGEMENT DECISION MAKING 3
- Coordination—the Decision-Making Function, 4
- Risk and Uncertainty, 5
- Sequential Decisions, 16
- Areas of Management Uncertainty, 18
- Summary and Conclusions, 21

2 FORECASTING METHODS 26
- Naive Methods, 26
- Lead-Lag Series and Pressure Indexes, 33
- Opinion Polling, 35
- Econometrics, 39
- Choice of a Forecasting Method, 47

3 ECONOMIC MEASUREMENT 51
- Measurement Methods, 51
- Graphic Multiple Correlation, 58
- Simple and Multiple Relations, 70
- Correlation Analysis in Forecasting, 75

PART II. ADJUSTMENT TO UNCERTAINTY

4 PROFIT MANAGEMENT 87
- Profit Theories, 87
- Unresolved Considerations, 91
- Profit Measurement, 94
- Profit Forecasting—Control, 109
- Nature and Dynamics of Profits, 117
- A Note on the Theory of Profit Maximization, 120
- Profit Policies, 124

5 DEMAND ANALYSIS—SALES FORECASTING 135
- Analytical Framework for Demand Measurement, 135
- Demand Determinants—Elasticity, 139
- Demand for Consumer Nondurable Goods, 161
- Demand for Consumer Durable Goods, 173
- Demand for Capital Goods, 189

6 PRODUCTION MANAGEMENT 202
  Production Functions: Simple Relations, 202
  Production Functions: Multiple Relation, 210
  Product-Line Policy, 219
  Operations Research—Linear Programming, 226

7 COST ANALYSIS 233
  Nature and Theory of Cost, 233
  Cost Measurement, 248
  Advertising Costs—Budgeting, 268
  A Note on Distribution Costs, 273

8 PRICING: PRACTICES AND POLICIES 279
  Pricing Concepts and Marketing Policies, 279
  Pricing Methods, 291
  Product-Line Pricing, 300
  Differential Pricing, 306

9 COMPETITION AND CONTROL 322
  The Antitrust Laws, 322
  Areas of Uncertainty, 329
  Measurement of Economic Concentration, 362

10 CAPITAL MANAGEMENT 368
  Administrative Aspects of Capital Management, 368
  Forward Planning of Capital Expenditures, 372
  Mainsprings and Problem Areas of Capital Management, 376
  Determining the Measuring Stick, 379
  Summary, 395

11 CAPITAL MANAGEMENT (CONTINUED) 398
  Establishing the Acceptance Criterion, 398
  Capital Cost Patterns, 413
  Corporate Capital Structures, 420
  Implications for Managerial Decision Making, 422
  Conclusion, 436

SUPPLEMENTARY PROBLEMS 440

INDEXES

AUTHOR INDEX 449

SUBJECT INDEX 451

PART I

Uncertainty and Prediction

Business economics, which is the subject matter of this book, may be defined as the integration of economic theory with business practice for the purpose of facilitating decision making and forward planning by management. With this in mind, the three chapters that follow provide a conceptual orientation to this branch of applied science. The underlying theme throughout is that knowledge of the future is uncertain, and yet executives must make decisions now and formulate plans for the future in the face of this uncertainty. The need for making forecasts, therefore, becomes evident, and the framework of uncertainty in which managers find themselves, combined with the prediction techniques available for reducing that uncertainty, constitute the area to which we now turn our attention.

Chapter 1

THE UNCERTAINTY FRAMEWORK OF MANAGEMENT DECISION MAKING

On August 2, 1954, a curious article appeared in *Life* Magazine. Written by Branch Rickey, one of baseball's most brilliant managers, the article dealt with the subject of baseball as a science. To the general reader it probably conveyed the impression of being a lesson in mathematics. But to the more thoughtful reader it posed the interesting question of whether the game of baseball might not someday be played on an electronic computer and the outcome predicted before the players ever get onto the field.

Branch Rickey wanted to know if a team's strength and weaknesses could be measured so that its efficiency might be improved. He gathered statistics and took them to mathematicians, who, upon analyzing the data, established that a team's standing could be closely approximated by taking the difference between its average runs over a season and the average runs of its opponents. With this as a guide, Rickey proceeded to set down what he believed to be the most important of the various factors that determine the scoring of runs. After considerable thought and research, he arrived at a mathematical formula for measuring a team's efficiency. With a few slight changes in symbols, and because of its seemingly formidable nature, it is presented in the footnote below where it may readily be overlooked by those who prefer.[1]

At this point one may ask, what do baseball and business have in common, other than their initials, and what is the purpose of starting a book in the economics of business management with a short discourse on the

---

[1] Let T_e denote team efficiency; H = number of hits; B_b = number of bases on balls; A_b = times at bat; H_p = hit by pitcher; T_b = total of bases; R = number of runs; H_b = hits by batsmen; E_r = earned runs; S = strikeouts; and F = fielding. Then

$$T_e = \left[\frac{H + B_b + H_p}{A_b + B_b + H_p} + \frac{3(T_b - H)}{4A_b} + \frac{R}{H + B_b + H_p}\right] - \left[\frac{H}{A_b} + \frac{B_b + H_b}{A_b + B_b + H_b} + \frac{E_r}{H + B_b + H_b} - \frac{S}{8(A_b + B_b + H_b)}\right]$$

where the first set of brackets represents the offense, and the second set, the defense.

science of competitive sports? The answer will become evident in the following sections and in later chapters as the coordinative or decision-making function of management is explained and illustrated.

COORDINATION—THE DECISION-MAKING FUNCTION

The functions of managers, whether baseball managers or business managers, may be classified for purposes of analysis into two distinct levels of activity: one of these is *coordination;* the other is *supervision.* The coordinative function is that of decision making—the process of selecting one action from two or more alternative courses of action. The need for this function arises in any type of environment where the future is uncertain, and yet decisions must be made and plans formulated by someone (or some group) on the basis of his (or its) expectations of the future. The other phase of management, that of supervision, involves the fulfillment of plans already established, and hence requires little, if any, coordination of a decision-making nature. It is management in the coordinative sense that will occupy the core of our attention throughout this book.

As in baseball, so too in business, the coordinator (rather than mere supervisor) of the firm's resources must continually choose between alternatives. Problems of choice arise because resources such as capital, land, labor, and management are limited and can be employed in alternative uses. The executive function from the coordinative standpoint thus becomes one of making choices or decisions that will provide the most efficient means of attaining a desired end, whether the end be the preservation of the *status quo* for the firm with respect to its competitors, or the ultimate one of gaining a monopoly in a particular market, or perhaps an intermediate one of profit maximization. But regardless of the goal, the fact is that business managers make decisions in a realm of uncertainty. If knowledge of the future were perfect, plans could be formulated without error and hence without need for subsequent revision. In most instances, however, the time involved precludes perfect knowledge of the future. Plans are made at one point in time that are based on current knowledge and involve current decisions, in anticipation of a result that will be forthcoming at some future point in time. As more facts become known, plans may have to be revised and a new course of action adopted if the desired objective is to be attained. Managers are thus engaged in the continuous process of charting a course of action into a hazy horizon.

What is the nature of this uncertainty that surrounds all management decision making where knowledge and prediction are imperfect? The answer to this is given in the following sections. What guides and tools are available to management for making choices and fulfilling plans in the face of uncertainty? The answer to this is given in the following chapters.

RISK AND UNCERTAINTY

The most pervasive characteristic of managerial decision making is imperfect knowledge of the future. Rarely do the firm's executives have complete information as to future sales, costs, and profits. They must continually estimate not only buyers' wants but also future materials, prices, wages, and productivity. Such estimates are essential if plans are to be made for the future and if operations are to be carried on in an efficient and profitable manner.

In making decisions and in formulating plans, business managers are confronted with two types of outcomes: *risk*, and *uncertainty*. Businessmen are prone to think of all types of outcomes that may result in losses as risks, but there are some technical distinctions between these two concepts that are fundamental for planning purposes and hence for practical problems to be discussed in later chapters.

Risk—Objective Prediction

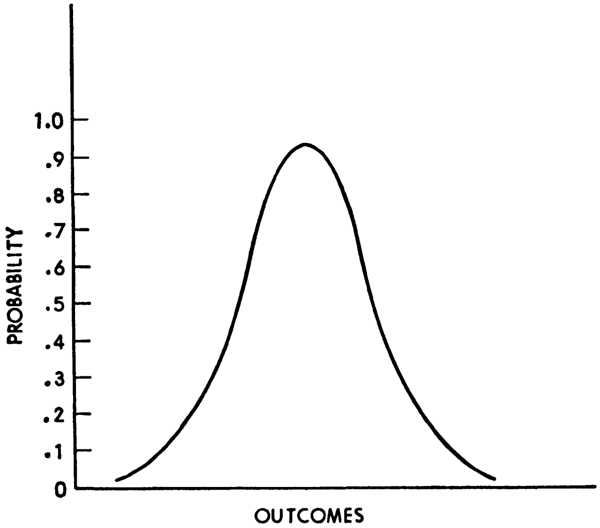
Risk may be defined in a business sense as the quantitative measurement of an outcome, such as a loss or a gain, in a manner such that the probability of the outcome can be predicted. The central ideas in the concept of risk are thus measurement and prediction, and the purpose of prediction is to estimate the likelihood of an eventuality or contingency. There are two methods that can be used in arriving at a probability measure of risk. One of these is by *a priori* deduction; the other is by *empirical measurement*. Both methods provide the information needed in making predictions.

In the *a priori* method, management can compute with certainty the probability of an outcome without the necessity of relying on past experience. Deductions are made on the basis of assumed principles, provided that the characteristics of the eventuality are known in advance. Thus, it is not necessary to toss a coin a large number of times in order to discover that the relative frequency of a heads (or tails) approaches $1/2$, or one out of every two tosses. Likewise, it is not necessary to make a continuous drawing of cards from a poker deck containing 52 cards in order to conclude that the probability of drawing any particular card is $1/52$. And with continuous rolls of a perfect die, it can be predicted with certainty that any given number, say 4, will turn up 1 out of 6 times, so the probability can be written as $1/6$, or 0.17. Probability statements such as these are not intended to predict a particular outcome. They merely state that in a large number of cases, this is the only outcome that will be realized with certainty. It follows, therefore, that the habitual gambler who entertains himself with *organized* games of chance is faced with risks, not with uncertainty, and the only thing that is certain is that he must lose over time.

Business managers rarely encounter practical problems involving *a priori* probability. This method of prediction is useful mainly in deriving and illustrating theoretical concepts. The more common method of predicting outcomes is by statistical or empirical measurement, because the results are based on actual experiences recorded in the form of past data. From a practical standpoint, the use of historical data for predicting the future assumes that past performances were typical and will continue in the future. In a stricter sense, this means that in order to establish a probability, the number of cases or observations must be large enough to exhibit stability, they must be repeated in the population or universe, and they must be independent.[2] If the assumption that the data are typical is a valid one, the statistical probability can then be computed and the likelihood of the outcome can be classified as a risk. Thus, insurance companies predict with a high degree of certainty the probability of deaths, accidents, fire losses, etc. Though they cannot establish the probability that a particular individual will die or that a particular house will burn, they can predict, with small error, how many people will die next year or how many houses out of a given number will burn.

FIGURE 1-1

PROBABILITY DISTRIBUTION OF OUTCOMES INVOLVING RISK



For eventualities or outcomes that involve risks, a primary task of business managers is to develop methods that will enable them to calculate and subsequently minimize the risks inherent in a prediction problem. The method used to accomplish this is to construct a frequency (probability) distribution of outcomes, as in Figure 1-1. Since the proportion of outcomes can never be negative, and since it can also never exceed 100 per cent or a relative frequency of 1.00, a probability must always lie between 0 and 1. A probability of 0 means that the event is extremely unlikely; a probability of 1 means that the event is likely to happen all of the time. Values in between denote degrees of certainty. On the horizontal axis the outcomes under consideration are scaled off. A frequency or probability distribution of outcomes is then plotted, which in Figure 1-1 is in the form of a normal curve. The characteristics (parameters) of the distribution should also be established for purposes of analysis. That is, a measure of central tendency is needed, such as the mean, median, or mode, to describe the typical size of the distribution; a measure of dispersion, such as the standard deviation, variance, or moments, to establish the scatter; a measure of skewness to denote the degree of symmetry; and a measure of kurtosis to describe the peakedness of the distribution. These measures can be established with an empirical probability of 1 (certainty) for the particular distribution. Risk would be present when the outcomes can be predicted over a period of years, in terms of these measures, and also the number of years in which the outcomes will fall.

---

[2] Independency means that the observations are distributed in the manner of a stochastic variable, i.e., at random.

What are the decision implications of risks, and of what significance are they in affecting management's role as a coordinator of the firm's resources? In its capacity as a decision maker and planner, the managerial function is essentially forward looking in nature. Plans are made in the present based on expectations of the future. Since it is a characteristic of pure or objective risk that the parameters (mean, variance, skewness, etc.) of the frequency distribution of outcomes can be predicted with certainty, the expected losses or gains can be incorporated in advance into the firm's cost structure. This is true whether the risk is of an intrafirm or interfirm nature; both of these types raise no problems for planning purposes, as explained below.

1. *Intrafirm risk* occurs when management can establish the probability of loss because the number of occurrences within the firm is large enough and hence predictable with a high degree of certainty. For example, a factory may experience a loss of 2 machine-hours out of every 100 machine-hours due to equipment breakdown. It might seem that this type of loss should be insured, but such is not the case. In reckoning profits, the cost of the production lost can be added to the cost of the production resulting from the remaining 98 machine-hours, and the profit rate will be altered accordingly by the revision in the cost structure. In other words, where the mean expected loss for the company can be predicted for the coming period, the loss can be treated as a cost of doing business and no insurance to cover such loss is even necessary. Therefore, since the future cost of the loss can be planned in advance of the company's fiscal period, intrafirm risk presents no problem for planning purposes.[3]

---

[3] Small-loan companies expect a certain percentage of defaults, banks charge off regularly a portion of their loans, and many companies have attempted to insti-

2. *Interfirm risk* occurs when the number of observations or cases is not large enough within any one firm for management to predict the loss with even near certainty. However, when many firms are considered, the number of observations becomes numerous enough to exhibit the necessary stability for prediction. Examples of such risks are losses caused by floods, storms, fires, etc. Since managers are unable to predict such losses for themselves, they are able to shift the burden of the risk to insurance companies whose function it is to establish the probability of such losses based on a large number of cases. Under such circumstances, the probability of loss for a specific firm cannot be predicted, but the probability of loss covering many firms can definitely be established with a small amount of error. It follows, therefore, that since the insured pays a (risk) premium for insurance, this can also be incorporated in the firm's cost structure for planning purposes. As with intrafirm risk, interfirm risk poses no problem of a planning nature. Being a predictable phenomenon, either type of risk can be treated as a future cost and, therefore, need have no bearing on management's role as a decision maker and planner. Uncertainty, on the other hand, is a subjective concept, and it is here that the real challenge to management manifests itself.

Uncertainty—Subjective Prediction

Like risk, uncertainty is also forward looking in nature, but unlike the former, it is not objective and does not assume perfect knowledge of the future. Uncertainty is a subjective phenomenon; no two individuals will view an event and necessarily formulate the same quantitative opinion. This is due to a lack of sufficient historical data on which to base a probability estimate, which in turn is caused by rapid changes in the structural (fundamental) variables that determine each economic environment. In other words, the observations are not repeated often enough to establish a probability figure based on repeated, homogeneous trials, as in the case of risk. Managers, therefore, must make decisions in an environment of incomplete knowledge, which they do by forming mental visions of the future that cannot be verified in any quantitative manner. It follows from this that uncertainty is not insurable, and cannot be integrated within the firm's cost structure as can risk. The parameters of the probability distribution cannot be established empirically, because all predictions are subjective and within the framework of each manager's own anticipations of the future. At best, subjective probabilities can be assigned to these anticipated outcomes, but the distribution of expectations resulting therefrom cannot be established with objective certainty. The following examination of the various types of uncertainty will help clarify these concepts.

---

tute self-insurance programs for various kinds of risk to which they are subject and feel they can prepare themselves against through proper reserve accounting.

Types of Uncertainty

The concept of uncertainty can be defined further by recognizing the different ways in which managers may view the uncertain outcome of an event. But first a clear statement of the problem as it confronts the decision maker may be helpful. Businessmen make decisions and formulate plans during the present time period (t_1) in anticipation of significant events that they expect will occur at a stated future time period (t_2). These decisions are made under conditions of imperfect knowledge as to the future. If the significant event being anticipated (such as a sales forecast) is realized, then the plans made in order to fulfill the prediction (e.g., setting up production schedules, capital requirements, etc.) will turn out to have been correct, and the firm will have made its full (equilibrium) adjustment to uncertainty. But if the significant event is not realized as anticipated, the original plans based on the expectation will be in error, and the plans will have to be revised as the firm approaches t_2. A significant event is thus an outcome or an occurrence which, if foreseen perfectly, would have influenced a particular decision and the formulation of a particular plan. It is for this reason that problems involving objective risk do not involve decision, since risk outcomes can be predicted with certainty and their consequent losses or gains can be planned for in advance. But for problems involving uncertainty, the keynote is subjectivism, and here decisions can only be made on the basis of anticipated outcomes.

With these considerations in mind, the nature of uncertainty can be further analyzed by noting three different ways in which managers may view the uncertain outcome of an event. They may view it with (1) subjective certainty, (2) subjective risk, or (3) pure or subjective uncertainty.

1. *Subjective certainty* is a type of uncertainty in which the business manager foresees only one possible outcome in the future period. For example, a manager in planning his production on the basis of a sales forecast might expect a particular sales figure for the coming year. He is subjectively "certain" that any other sales volume will not occur, and hence assigns no "weight" to any values higher or lower than this single amount. Compared to the other two types of uncertainty outlined below, this type is relatively rare. Illustrations of it are found, however, in all situations where prices are regulated by law, such as price control during war time, public utility rate regulations, and the like. Managers in such circumstances can plan with subjective certainty that prices will not be higher or lower than the legal level. Strictly speaking, however, it is a contradiction in ideas to regard subjective certainty as a type of uncertainty. For when a decision maker entertains only one possible outcome with subjective certainty, i.e., harbors a single-valued expectation as explained above, he is saying in effect that he is certain rather than uncer-

tain. True uncertainty would require that the manager view not one, but a range of outcomes as being possible, or that his expectations be multivalued rather than single valued, and that these eventualities be "weighted" in some sense. This is characteristic of the remaining two types of uncertainty discussed next.

2. *Subjective risk* exists when managers, knowing that the future is uncertain, accept a range of outcomes, rather than a single outcome, as possible. This anticipated range of outcomes may be in the form of a single probability distribution, but the parameters of the distribution cannot be measured in an objective or empirical manner. Hence, the manager who views a distribution of outcomes with subjective risk is saying in effect that he anticipates the distribution as a whole with "certainty," but he cannot anticipate any particular outcome within the distribution with certainty. To illustrate the concept, the manager may frame his distribution in an ordinal (ranking) sense by saying that sales will range between $405 million and $415 million. This is the distribution as a whole which he anticipates with subjective certainty. He would then rank these outcomes in order of certainty, starting, say, with his most confident expectation and decreasing to the least confident or most uncertain one. Thus his distribution might appear as in Table 1–1.

TABLE 1–1

"SUBJECTIVE RISK" DISTRIBUTION. ORDINAL

$410 million	(most confident)
$412 million	
$408 million	(fairly confident)
$405 million	
$415 million	(least confident)

When the distribution is framed in a cardinal sense, the manager might assign subjective probability values to each possible outcome. The result would then be a frequency or probability distribution with the subjective probabilities acting as weights. A distribution of outcomes arranged in this manner would be held with "certainty," and might appear as shown in Table 1–2.

TABLE 1–2

"SUBJECTIVE RISK" DISTRIBUTION: CARDINAL

Expected Sales (Millions)	*Percentage Chance of Realization*	*Probability*
$405	10%	0.1
408	20	0.2
410	40	0.4
412	20	0.2
415	10	0.1
	100%	1.0

A number of expectation studies along these lines have been conducted among farmers, investors, businessmen, and consumers for various types of forecasts, and it has been found that a significant number actually do view the future at least in the ordinal sense outlined above.[4] The cardinal type of distribution, though it was rarely encountered unless the questionnaire was couched in these terms, may be regarded as a more refined method of attempting to estimate uncertainty.[5]

3. *Pure* or *subjective uncertainty* in contrast to the previous types discussed above, represents the most complex aspect of uncertainty. Here the manager views not a single distribution of outcomes as possible, but multiple distributions which, in turn, are predicated on his probability expectations of structural changes in the economic environment. For example, the managers of a company in a purely competitive industry, in planning production for a future time period, may want to formulate production schedules based on their expectations of the industry's total output. The managers of the firm recognize that the industry's sales will depend in large part on whether the economy as a whole is in a recession, remains relatively normal, or is in a prosperity period. Accordingly, they agree on a probability of .2, .5, and .3, respectively, for each of these possible structural changes in the economic environment. In other words, they believe that there is a 20 per cent chance for recession, a 50 per cent chance for normality or *status quo*, and a 30 per cent chance for prosperity, thus making a total of 100 per cent or a probability of 1. For each of these possible structural outcomes, they establish a range of possible outputs with a subjective probability value attached to each. The result is a probability distribution of probability distributions as shown in Table 1–3.

The managers of the firm now have two choices confronting them: (1) They can adopt a "simple" method of calculation whereby they accept the normal structural environment as most probable ($P = .5$) and, on this basis, plan their own production schedule with the expectation that the industry will sell 480 thousand units, since this is the output that is most probable ($P = .4$) in this distribution; (2) they can use a "weighted" method whereby they merge the distributions into a single distribution by multiplying each of the three structural probabilities ($P = .2, .5, .3$) by the

---

[4] Economists have also been made the subject of such studies. At two recent conventions of the American Economic Association, seventy economists were interviewed in an attempt to establish ordinal and cardinal distributions of expected GNP outcomes. The results followed a very similar pattern to those illustrated above, as well as to the "pure uncertainty" examples shown below.

[5] The assignment of probability values or weights is based on the assumption that the variable being measured is randomly distributed. For sales estimates, this assumption is probably false for most firms. For price estimates, on the other hand, it might be applicable for firms in pure competition since each has no control over market price. In any case, the purpose of this discussion is to impress upon the average (nonmathematical) reader the role of probability as the basis for decision making, even at the possible expense of what we would regard at this stage as a minor error at most. Somewhat more significant statistical considerations are discussed in later chapters.

TABLE 1-3

"PURE UNCERTAINTY" DISTRIBUTIONS

RECESSION ($P = 2$)		NORMAL ($P = .5$)		PROSPERITY ($P = .3$)	
Industry Output (Thousands)	Probability	Industry Output (Thousands)	Probability	Industry Output (Thousands)	Probability
450	0.2	450	0.0	450	0.0
460	0.4	460	0.0	460	0.0
470	0.3	470	0.1	470	0.0
480	0.1	480	0.4	480	0.0
490	0.0	490	0.3	490	0.2
500	0.0	500	0.2	500	0.6
510	0.0	510	0.0	510	0.1
520	0.0	520	0.0	520	0.1
	1.0		1.0		1.0

probabilities listed under it. These products could then be summed to give a compounded or total probability distribution. For the distribution shown above, the multiplication procedure is illustrated in Table 1-4, with the resulting total (weighted) distribution given at the right.

The total probability distribution at the right provides management with a conceptual basis for a plan of action. Note that the most (subjective) probable industry output is 500 ($P = .28$), as compared to the previous "simple" case where the expected output would have been 480 ($P = .4$) had the "normal" structural outcome been chosen as most probable. Clearly, the formulation of a proper plan by management can be greatly affected by the choice of rational calculation that it adopts. In effect, the compounding method "weights" the estimates of outcomes in accordance with their relative importance to the distributions as a whole, and is therefore preferable to the "simple" method outlined previously.[6]

Do businessmen actually formulate their expectations in the refined manner discussed above? Most of them do not. However, they do scan business magazines, outlook reports, and any other sources that may be available in order to frame, at least in an ordinal sense, their expectations

[6] A. G. Hart has developed these concepts in greater detail. See his *Anticipations, Uncertainty, and Dynamic Planning*, chap. 4, and "Risk, Uncertainty, and the Unprofitability of Compounding Probabilities," in *Studies in Mathematical Economics and Econometrics*. Hart points out that knowledge is lost through compounding because the total probability distribution has only one sum, and yet it may represent the sum of any number of component distributions. Thus, the total distribution does not disclose whether the sum represents (1) many component distributions with a small dispersion each, (2) few component distributions with a wide dispersion each, or (3) many component similar distributions with a large dispersion each. Thus there are "degrees of uncertainty" which may have a significant bearing for planning purposes. (See also E. O. Heady, *Economics of Agricultural Production and Resource Use*, chap. 15, for discussions of these concepts with reference to agriculture, where much of the theory of uncertainty originated.

TABLE 1-4

COMPOUNDED "PURE UNCERTAINTY" DISTRIBUTIONS

Industry Output	Recession (Probability)	Normal (Probability)	Prosperity (Probability)	Total (Probability)
450....	0.04	0.00	0.00	0.04
460. . .	0.08	0.00	0.00	0.08
470....	0.06	0.05	0.00	0.11
480	0.02	0.20	0.00	0.22
490....	0.00	0.15	0.06	0.21
500	0.00	0.10	0.18	0.28
510. .	0.00	0.00	0.03	0.03
520......	0.00	0.00	0.03	0.03
	0.20	0.50	0.30	1.00

as to the future. And consumer surveys, as a method of forecasting (discussed in the next chapter) are based in large part on these ideas. Regardless of the method employed, decision making, if it is to be a science, requires a logical and systematic analysis of expectations. Preferably, these expectations should be quantified as much as possible, for business plans are formulated on the basis of such results. Predictions, whether they are based on hunches, ordinal expectations, or refined cardinal expectations, must be made whenever knowledge of the future is less than perfect. It should be useful, therefore, to examine the problems of prediction by way of the degree of confidence, or degree of uncertainty, that exists when attempts are made to forecast the future.

Degree of Uncertainty

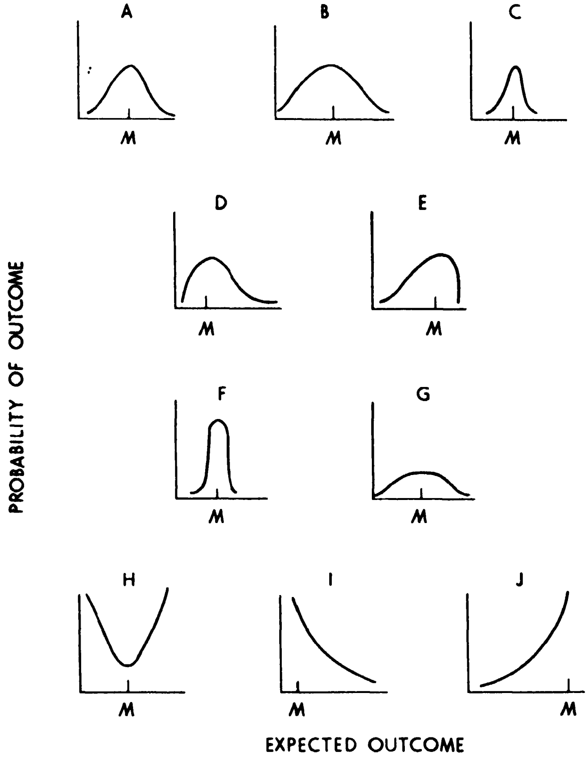
Since expectations are subjective in nature, there will be "degrees of uncertainty" on the part of managers. Two businessmen may view the same event, but each will establish his own expectations with greater or lesser confidence than the other. The probability or frequency distributions relating to future events are not objective or empirical but only subjective or imagined by each individual. Therefore, an examination of the various ways in which the "degree of uncertainty" may be represented graphically should be of value for purposes of analysis.

Degrees of uncertainty may be illustrated by the set of probability or frequency distributions shown in Figure 1-2. Expected outcomes, such as a forecast of sales, costs, GNP, etc., are plotted horizontally, while the subjective probability of the outcome is measured off vertically.

In the top panel of Figure 1-2 there are three symmetrical distributions, but only A is of normal form. In contrast with B and C (all of which are plotted in the same units), A represents greater uncertainty than C. This is due to the variance or spread of the distributions. If M is taken as the modal or most probable outcome, the greatest variation in expectations (or greatest uncertainty) occurs in B because it has the widest spread; the least variation or uncertainty is in C because it has the least spread. Thus, as an indication of the degree of uncertainty, a measure of the *dispersion* of expectations such as the range, the standard deviation, or the variance could be used. The degree of uncertainty could be said to vary directly with the dispersion: a dispersion of zero would mean perfect certainty or a single-valued expectation; a larger dispersion would indicate greater uncertainty or multivalued expectations.

FIGURE 1-2

SUBJECTIVE PROBABILITY DISTRIBUTIONS



In Figure 1–2 D and E illustrate the significance of *skewness*. Compare these curves with the normal curve of A, where the expected outcomes are arranged symmetrically about the modal outcome. In A, there is an even chance that a deviation from the most probable (modal) outcome will be greater or less than the mode. In D, a deviation from the mode *M* will most probably be in the direction of higher values, while in E the most probable deviation is toward lower values. Thus, a measure of skewness, since it describes the "lopsidedness" of the distribution, is a further indication of the degree of confidence or uncertainty.

A comparison of F with G reveals still further characteristics. The degree of peakedness of a distribution, called *kurtosis* when measured statistically, is yet another indication of uncertainty. In F the probability of the particular modal outcome *M* is greater than for any other distribution shown. In G, however, where the distribution is relatively "flat topped," the subjective probability of the modal outcome is only slightly greater than outcomes higher or lower. In F, the manager is "relatively certain" of the modal outcome; in G, his expectations cover a wide area of almost equal probabilities, and he would find it difficult to formulate plans based on the modal outcome or on outcomes higher or lower than M.

In Figure 1–2 H, I, and J represent a different pattern of distributions. The U-shaped distribution of H implies high, equal probabilities for outcomes whose values are either large or small, and a low probability for outcomes in between. The J curve of J shows a high probability of higher-valued outcomes, while the "reverse" J curve of I implies a high probability for the lower-valued outcomes. These latter two curves have statistical connotations similar to those of E and D, respectively.

It may be noted that, technically, there is some disagreement among economists as to whether the mode or the mean should be used as the "expected value" for prediction purposes. One group, composed of Hicks, Lange, and Fellner, contends that the mode is preferable as the most probable value because it is more realistic: a decision maker is unlikely to calculate the mean of a probability distribution whose shape may not be clear cut to begin with.[7] The other group, consisting of Pigou, Hart, and Tintner, favors the mean value because it is the theoretically correct one from which to compute the standard deviation as a measure of dispersion for the distribution.[8] In any event, it is beyond the scope of this book to go into the ramifications of these arguments. What concerns us most is the development of a conceptual orientation—a way of thinking about the nature of uncertainty—that forms a basis for scientific decision making and planning in business administration.

Conclusion

The basis of decision theory as outlined above takes place in a framework of uncertainty. The distributions of Figure 1–2 are purely subjective with each manager and should not be viewed in the same statistical sense as those derived on the basis of (objective) risk phenomena. The latter imply that the probability estimates are established by repeated and independent trials, whereas the uncertainty distributions allow no such advantage to the business manager. In reality, the funds available to the company at any given time are limited, and the opportunity to repeat a trial or even to modify a past decision in the light of new evidence may not always exist. Where the action is crucial, a wrong decision can sometimes spell either success or bankruptcy. Fortunately, for well-established firms, it more often involves a difference between profit or loss on the income statement for the accounting period. Probabilities, therefore, should not be viewed for decision purposes as long-run frequency ratios since economic events rarely repeat themselves in a homogeneous manner. It is the *degree of uncertainty* that is relevant for decision making, and probability should be looked upon as a connecting link between the evidence available and the outcome being considered, with neither being necessarily measurable. As the available evidence becomes larger, the "weighted" probability of a particular event relative to others increases, and the degree of uncertainty thus diminishes accordingly.

---

[7] See J. R. Hicks, *Value and Capital*, 2d ed., p. 125; O. Lange, *Price Flexibility and Employment*, p. 30; and W. J. Fellner, *Monetary Policies and Full Employment*, p. 152 ff.

[8] See A. C. Pigou, *Economics of Welfare*, 4th ed., App. I; A. G. Hart, *op. cit.*, p. 52; G. Tintner, "A Contribution to the Non-Static Theory of Production," *Studies in Mathematical Economics and Econometrics*, p. 99.

SEQUENTIAL DECISIONS

The discussion of decision making thus far has stressed a rational approach to managerial action within a framework of uncertainty. Managers are confronted with choices to be made because knowledge of the future is imperfect. If future eventualities could be known with certainty, there would be no need for decision theory; economic plans would be formulated in a timeless, once-and-for-all vein, and expectations of future changes would be nonexistent. Only management in the supervisory sense would be needed. But in the real world, operating decisions must be made with the forward-looking recognition that plans as originally formulated may not be realized. What are the aspects of operational decisions, therefore, that managers must encounter in establishing their plans? Though the answer to this question is mathematical in nature, a brief look at the conceptual side of this problem will help indicate the types or levels of decision making that actually exist in a world of uncertainty.

Many decisions involve situations in which past experiences are either nonexistent or extremely scarce. A new type of advertising campaign or the introduction of a new product may require a plan of action for which an analysis of the past results of related activities is not available. The possible courses of action based on historical experiences are not at all evident, but a choice must nevertheless be made and a plan formulated. Moreover, not one, but a *series* or *sequence* of decisions may have to be made in order to achieve a desired goal. These sequential decisions may be simultaneous at the same level of management (i.e., horizontal sequences) and be revised as new information becomes available, or they may be strategic decisions made at the staff level and be transmitted down to where they are utilized as tactical procedures in the basic operations of the firm. Sequential decisions are thus a recognition that a multiplicity of changing factors may enter into the formulation of a plan, rather than a fixed set of circumstances to be reckoned with in a once-and-for-all or terminal manner.

Levels of Sequential Decisions

In a statistical sense, sequential decisions may be either of a "narrow" or "broad" type.[9] When they are narrow in nature, observations are made in discrete units (one at a time) and, after each observation, on the basis of the information then known, management has the choice of either making a final decision and formulating a plan, or making further observations before deciding on a final plan. The alternatives, therefore, are either to make a decision based on the information already known, or to gather further information along the same line until it is felt that enough is already known to make a terminal decision. "Broad" sequentials, on the other hand, pose two different kinds of alternatives: either to discard all past observations and establish new ones, or to conduct an experiment in order to determine the efficiency of the collection process. Both these types of sequentials are employed in management decision making.

The basic problem in narrow sequentials is to know when enough data (observations) are available to make a decision. This requires that the necessary prior information be gathered, and that the costs and consequences be examined. For example, in conducting a marketing research study on consumer preferences, it would be necessary to establish (1) the cost of sampling, (2) the cost of wrong answers due to misrepresentations on the part of respondents in answering the questionnaire, and (3) the cost of rejecting all the replies and taking a 100 per cent sample instead. Before a decision can be made, it would also be necessary to establish the prior probabilities, based on what is already known about the products from past experiences as to consumer preferences. When this information is known, management can proceed in a rational way to analyze a problem in order to arrive at an appropriate decision.

Until now, most of the research done on narrow sequentials in the field of business administration has been confined to sequential sampling of industrial products purchased for business use. The armed services have also employed narrow sequential decision techniques in the purchase of munitions and other goods where, for example, the sampling is destructive and yet economy in sampling is not to be sacrificed for accuracy. It seems likely that these methods will have increasing applications to managerial decision making in the years to come.

In contrast with the narrow type of sequential decisions, where the analysis proceeds in stages and a decision must be made at each stage as to whether to go ahead or make a terminal decision, broad sequentials deal with decision problems that are more comprehensive in nature. For example, to measure the effect of a new package design on total sales, management would have to make a series of preliminary decisions before arriving at a terminal decision as to whether the new package should be adopted or the old one should be retained. The sequence of decisions might involve: (1) an examination of the packaging policies of competitors, (2) the setting up of a planned experiment in which the effectiveness of various package designs is tested and measured, and (3) isolating those package designs that are most promising and testing them still further. It may happen that a stated sequence of decisions such as these results in an impasse somewhere along the line because certain required information needed in the series is unobtainable. When this occurs, the data may have to be discarded and a new sequence established, or the problem may have to be reconstructed and attacked in an entirely different manner. Either alternative is possible, depending on the type of problems at hand and the information already known.

---

[9] *Cf.* I. D. J. Bross, *Design for Decision*, chap. 8.

Uncertainty Effects

The effect of uncertainty is to limit the size of the firm by limiting management at all levels of decision making. Where knowledge is imperfect and uncertainty prevails, managerial responsibilities become greater. Managers have more decisions to make within a given time period; therefore, decisions become less perfect as supporting knowledge is reduced. Even the addition of more managers must eventually be limited since, ultimately, strategic decisions still must pass through a central management group. This is not to say that the size of any particular plant is subject to such limitational effects; only the growth of the firm is constrained as uncertainty becomes greater, decisions become more imperfect, and "diminishing returns to management" set in.

The significance of this with respect to sequential decisions should be noted. As the number of decisions to be made increases, less time is available for gathering information relevant to any particular decision, and the result is a less perfect prediction of outcomes. Conceivably, if the method of sequential decision making can be generalized sufficiently to include most kinds of managerial problems, the limiting effects of uncertainty will be reduced (although not necessarily eliminated). Sequential decision making, in other words, offers a more efficient method of arriving at terminal decisions by systematizing procedures with respect to the gathering and testing of information. In effect, therefore, it enables managers to make a greater number of correct decisions within the same allotted time, or the same number of correct decisions in less time.

AREAS OF MANAGEMENT UNCERTAINTY

The executives of a business organization are its ultimate decision makers, and their ability to make correct decisions will determine the future course and well-being of the firm. Since decisions must be made in the face of uncertainty, the obvious question is: What are the common areas of uncertainty that confront top management in its role as coordinator of the firm's resources? Intuition and the experience gained from personal observation alone would lead us to believe that the number of such uncertainty areas, if narrowed down, would be quite large. The fact that numerous corporations each spend many thousands of dollars annually on business research which, in the final analysis, is aimed at delineating and measuring uncertainty situations so as to facilitate improved decision making, is further indication of its complex magnitude. Instead of narrowing these uncertainty areas down, which will be a chief function in Part II of this book, it is sufficient to state them now in terms of broad classes of problems. This will serve to outline the direction in which we are headed and thus provide a better perspective of the subject matter as a whole.

Profit Uncertainty

Most business firms are organized for the purpose of making profits, and in the long run profits are the chief measure of success. From the management viewpoint, the more relevant considerations involve an understanding of the nature and causes of profit, the methods of profit measurement, the choice of an appropriate profit policy, and the alternative techniques of profit planning. Profit uncertainty exists because of variations in costs and revenues, which in turn are conditioned by factors both internal and external to the company. If knowledge of the future were perfect, managers could make smooth and immediate adjustments to economic changes, and profit variability would be minimized. In a world of uncertainty, however, expectations are not always realized, and hence profit management and planning constitute the essence and ultimate purpose of managerial control. These problems are first posed and discussed in Chapter 4.

Demand Uncertainty

A business firm is an economic organism which transforms productive resources into goods that are to be sold in a market. A major portion of managerial decision making depends on accurate estimates of demand. Before production schedules can be set up and resources employed, a forecast of future sales is essential. Once the forecast is made and plans for the future are formulated, a knowledge of the demand for its products can serve as a guide to management for maintaining or strengthening market position and enlarging profits. Demand analysis, which encompasses both demand forecasting and measurement, is essential for business planning and hence occupies a strategic role in managerial economics. Its essentials are presented in Chapter 5.

Production Uncertainty

The challenge of organizing the firm's production process so that maximum efficiency can be attained from given resources poses a further decision area for management. The choices to be made that are of a decision nature involve essentially (a) how resources should be allocated by management between different products or production methods at a given point in time, and (b) how resources should be allocated to products over periods of time. The first is largely a static problem involving the hiring of productive resources; the second is a dynamic one concerned not only with factor hire, but with the establishment of a product-line policy in accordance with available company resources and changing demand conditions. Both problem areas are taken up in Chapter 6.

Cost Uncertainty

A study of economic costs combined with data drawn from the firm's accounting records can yield significant cost estimates that are useful for management decisions. The uncertainties that cause variations in costs must be recognized and allowed for if management is to arrive at cost estimates that are significant for planning purposes. Cost uncertainty exists because all of the factors determining costs are not always known or controllable. Discovering economic costs and being able to measure them is a necessary step for more effective profit planning, control, and often for sound pricing practices. These problems are treated in Chapter 7.

Pricing Uncertainty

The analysis of demand, production, and costs provides the basis for a study of pricing. The problems here involve an examination of alternative pricing policies and pricing structures for the purpose of minimizing the uncertainties inherent in establishing marketing and dstribution strategies. The means of accomplishing these tasks is through an integration of pricing theory and pricing research, attained in a practical manner through the construction of an operational base that provides management with a skeleton guide for judging the short- and long-run effectiveness of pricing decisions. These and related problems comprise the subject matter of Chapter 8.

Political-Economic Uncertainty

Management decision making frequently involves issues the legality of which are subject to the antimonopoly laws and possible scrutiny by the antitrust agencies. Executives, therefore, need at least a basic understanding of the leading antitrust issues and the recent trends followed by the courts in judging firms accused of violating the law. Among the chief topics of concern are the legal status of monopoly, conspiracy, patent licensing, trade-mark infringement, price discrimination, and resale price maintenance, not only by themselves, but also in terms of the philosophy of the antitrust agencies at the present time. These and similar topics fall under the heading of competition and its control and are discussed in Chapter 9.

Capital Uncertainty

Perhaps the most important area of uncertainty confronting management decision makers is that of investing stockholders' money. The need for performing this function properly gives rise to a number of problem areas relating to the scale of the firm, risk aversion, equity, and the credit market. Implicit also is the valuation of the company's resources and the procedures employed for compounding costs into the future and discounting revenues back to the present. These and a host of other problems comprise the subject matter of a relatively new field in business administration—that of capital management. Briefly, it deals with the administration, planning, and control of capital expenditures, or, in other words, the budgeting of capital under conditions of uncertainty. In view of its complexity and, in a broad sense, its over-all significance to management, it is given a more extended treatment than previous topics as will be seen in Chapters 10 and 11.

The seven classes of topics outlined briefly above represent most of the broad areas of uncertainty to which adjustments must be made by management as the firm's future course of activity is organized and planned. The procedures for adjusting to these uncertainty areas represent the subject matter of Part II. Before these topics can be treated, however, there are still certain considerations to be noted with respect to the existence of uncertainty. Basically, these considerations relate to the need for prediction and measurement as a means of minimizing the uncertainty inherent in most business situations. Before any plans can be formulated with respect to profit, demand, production, etc., management needs at its disposal a set of tools by which essential relationships can be quantitatively measured and the outcome of particular events may be numerically predicted. The general methods that can be used for accomplishing these ends are presented in the two chapters that follow. The application of these methods to specific areas of uncertainty will then be used to form much of the subject matter in the remaining portion of this book.

SUMMARY AND CONCLUSIONS

A discussion of uncertainty and the way in which it affects management decision making is not "theoretical and impractical." It is because managers are faced with uncertainty (imperfect knowledge of the future) that they must make decisions the outcomes of which are not known in advance; and it is because they lack perfect knowledge of the future that they must formulate plans which, due to unforeseen (unpredicted) contingencies, may not be realized. Prediction is thus an essential part of managerial economics because it is a necessary step in establishing plans for the future. In a world of perfect certainty where all future outcomes were known in advance, managers in the coordination sense as decision makers and planners would be unnecessary; only managers in the supervisory sense would be needed, for once the firm were established and resources committed, operations would take place in fulfillment and in accordance with an initial once-and-for-all plan.[10]

To be of practical effectiveness to executives, predictions must usually be formulated quantitatively rather than qualitatively. This means that for the major areas of uncertainty that confront managers, namely profit, demand, production, cost, pricing, competition, and capital uncertainty, prediction or measurement methods must be devised that will enable forecasts to be made in numerical terms. The techniques commonly employed for such purposes are classified and discussed in the following two chapters, and their application to the above uncertainty areas are illustrated throughout most of this book. Much of the study of business economics, it will be seen, involves the application of economy theory to practical business problems, formulated in a conceptually quantitative manner so that essential relationships may be measured and predictions made as a basis for policy development and planning. In any activity where knowledge of the future is imperfect, decisions are made in the present based on expectations of the future, and therefore prediction forms the connecting link between the known facts of today and the uncertain events of the future.

Economic Insurance

It is interesting to inquire as to why insurance against the uncertainty of losses has never been instituted in the business world. Enough has already been said to distinguish risk from uncertainty, and it should be clear that insurance can only be a response to the former rather than the latter. However, some further comments on the subject may be of interest at this point.

Fundamentally, there is no means known by which uncertainties can be isolated and generalized to the point where they may be classified as risks. Marketing researchers have yet to discover the economic laws (if any) that govern consumer behavior, and economists have yet to reduce to precise mathematical formulas the forces determining business cycles and the losses resulting therefrom. In other words, there are too many imponderables to take into account in arriving at an exact method of prediction. In a statistical sense, the determining factors are not randomly distributed, they are not homogeneous, and they are not quantifiable enough to establish results that can be predicted with known error. Furthermore, the creation of an adequate insurance program, if such were possible, would probably have to be based on long-run changes in prosperity and depression.

---

[10] "It is fascinating to contemplate political or social institutions in a perfectly certain world. Policemen would be needed only at the moment of crime, firemen only at the time of fire; laws could be more pointed to cover precisely the specific objectionable future occurrence, etc. The mere knowledge of future developments would not always be sufficient to prevent anti-social outbreaks, for some of them may not always be amenable to advance control." (S. Weintraub, *Price Theory,* p. 349.)

During prosperity periods when most companies are profitable and losses are at a minimum, the total amount of insurance premiums paid should exceed the total amount of losses incurred. Conversely, in depression periods losses are more general and should exceed the premium payments. Over the long run, therefore, the insurance company's profit during prosperity would presumably be balanced by its deficits during depression. But this would assume that the amplitude and frequency of business fluctuations could be forecast with accuracy, which they cannot, and this is precisely the problem at hand. Moreover, if such accurate forecasts of depressions were possible, insurance would be unnecessary; action could be taken instead to avoid the setback thereby eliminating the need for insurance.[11]

The role of uncertainty also manifests itself with respect to income and employment. Broadly, as certainty increases in industries because of improved human ability to predict, incomes there will tend to decline in the long run relative to the more uncertain fields of employment. At least this is true to the extent that individuals choose and train for occupations on the basis of their probability estimates of future earnings. The greater the uncertainties of the future in an occupational line, as indicated by the dispersion of earnings probabilities, the more attractive is the drive for security (in return for lower incomes) in the more certain fields of employment. Witness the fact that employees in civil service, banks, and public utilities generally earn less than their colleagues in the more uncertain though similar lines of activity in other industries. Admittedly, these monetary differentials may be somewhat compensated by nonmonetary factors (e.g., longer vacations, shorter hours, retirement benefits, etc.). But the fact remains that coal miners earn more than ditch diggers; window washers in skyscrapers command a higher wage than dishwashers in restaurants; and college professors earn less (but live longer) than corporation executives. The clash between uncertainty and security can manifest itself in many ways. The removal of uncertainty, if that were possible, would serve to modify the structure and pattern of resource allocation and income distribution, and would create a type of economy substantially different from that which currently prevails.

It is probably fortunate, all things considered, that insurance of the type discussed is nonexistent. When managers have less reason to fear the possibility of loss due to a wrong decision because such loss will be largely protected, the incentive for greater efficiency and improved planning is substantially weakened. Uncertainty has created a venturesome spirit as the essence of American capitalism and free enterprise, and it seems neither possible nor desirable that this be sacrificed for an organized scheme of profit protection.

---

[11] *Ibid.*, pp. 348–52.

BIBLIOGRAPHICAL NOTE

The original distinction between risk and uncertainty was developed in the pioneering but difficult work by Frank Knight, *Risk, Uncertainty and Profit,* which represents a landmark on the subject. Other more recent studies include A. G. Hart, *Anticipations, Uncertainty and Dynamic Planning,* and his "Risk, Uncertainty, and the Unprofitability of Compounding Probabilities," in *Studies in Mathematical Economics and Econometrics.* The latter is a short discussion of the probability compounding approach to pure or subjective uncertainty. Some further writings on the same general theme are K. J. Arrow, "Alternative Approaches to the Theory of Choice," *Econometrica,* Vol. 19; G. L. S. Shackle, *Expectations in Economics,* chaps. 1 and 2; and two articles developing the notion of "subjective risk" by G. Tintner, "The Theory of Production Under Non-Static Conditions," *Journal of Political Economy,* 1942, and "A Contribution to the Non-Static Theory of Choice," *Quarterly Journal of Economics,* 1942.

Less technical works include the following. For those desiring to strengthen their background in economic theory, there are good single chapter expositions of risk and uncertainty in W. J. Baumol, *Economic Dynamics;* S. Weintraub, *Price Theory;* and E. O. Heady, *Economics of Agricultural Production and Resource Use.* The first two works emphasize the translation of uncertainty concepts into the indifference curves of economic theory; the last presents an excellent comprehensive treatment of principles, but with applications entirely to farming. On a still less rigorous level is a light and humorous work by I. D. J. Bross, *Design for Decision,* which contains some excellent material on decision theory and related topics. As useful complements to the approach taken in this chapter, there is an article by R. K. Gaumnitz and Brownlee, "Mathematics for Decision Makers," *Harvard Business Review* (May–June, 1956), which, despite its possibly misleading title, is entirely nonmathematical and stresses the sort of education executives will need for better decision making; also, three sources emphasizing qualitative and sociological (rather than statistical) aspects of decision theory are: R. Tannenbaum, "Limitations on Decision-Making," *Journal of Business* (January, 1950); R. Owens, *Introduction to Business Policy,* chaps. 8–10; and M. H. Jones, *Executive Decision Making.*

QUESTIONS

1. What similarities do you see, if any, between the management of a baseball team and the management of a business enterprise?
2. An expert baseball player may know nothing about the physics of moving balls, i.e., ballistic science, and yet be successful because he practices the principles of ballistics instinctively. Can the same be said of a business manager with relation to economic theory?
3. In your own words, define (*a*) certainty, (*b*) risk, and (*c*) uncertainty.

4. (*a*) What conditions are necessary in order for an outcome to be classified as a risk? (*b*) Of what significance is this?
5. Distinguish between (*a*) intrafirm risk, and (*b*) interfirm risk.
6. Classify each of the following as an intrafirm or interfirm risk: (*a*) glassware and china breakage in a restaurant; (*b*) egg breakage on a dairy farm; (*c*) absenteeism in a plant; (*d*) "acts of God" (cite examples).
7. What characteristic do intrafirm and interfirm risk have in common from a planning standpoint?
8. The Treasury Department buys all gold offered to it at $35 an ounce. From the mining company's viewpoint, is this price a risk or an uncertainty, and if either, which specific kind of risk or uncertainty?
9. (*a*) How does subjective certainty differ from "subjective risk"? (*b*) Are these actual types of true uncertainty? (*c*) What is the difference between an ordinal and a cardinal distribution of "subjective risk"?
10. Why is "subjective risk" used in quotation marks?
11. From a farmer's standpoint, is the price of wheat a risk or an uncertainty? What about the sale of next year's Plymouths from Chrysler's standpoint?
12. What is pure or subjective uncertainty?
13. Here is an interesting experiment in pure or subjective uncertainty.

    It is likely that, at least to some extent, your grade in this course will depend upon the average grade of the class as a whole, being higher if the class average is high, and lower if the class average is low. At this stage in the course it is reasonable to assume that both your grade and the class average are pure or subjective uncertainties. Therefore, using the class average as your structural environment, consider the three possibilities that this average will be a below-passing grade, an average-passing grade, or an above-passing grade, and assign your own expected probability values to each, as illustrated in Table 1–3, p. 12. Then, under each structural probability, list the possible grade you may receive (e.g., *A*, *B*, *C*, *D*, or *E*) and assign your own probability value to each. Construct a pure uncertainty distribution as in Table 1–3, and then construct a compounded pure uncertainty distribution as in Table 1–4, p. 13. Is there a difference between your final results? Try this experiment at various stages in the course and note whether any changes occur as you gain more "knowledge" and your expectations become less uncertain as you approach your planning horizon, i.e., the end of the course.
14. (*a*) If knowledge of the future were perfect, would there be "degrees of uncertainty"? (*b*) What is the significance of Figure 1–2, p. 14?
15. "The effect of uncertainty is to limit the size of the firm by limiting management at all levels of decision making." True or false? Explain, with respect to sequential decision making.
16. What are the various areas of uncertainty to be examined in this book? Comment briefly upon each as to the emphasis and approach which will be adopted.
17. In the light of the "Summary and Conclusions" beginning on p. 21, do you think industrial progress, or economic development in general, would be faster or slower in a perfectly certain as compared to an uncertain world? Comment.

Chapter 2

FORECASTING METHODS

The existence of uncertainty and the consequent need for forecasting has already been recognized in the previous chapter. Virtually every business and economic decision rests upon some type of forecast of future conditions. Successful forecasting reduces the areas of uncertainty that surround decision making with respect to cost and profit budgeting, production and employment stabilization, inventory control, pricing, investment planning, and a host of other problems confronting business managers. What methods are available to managers for forecasting future business conditions both for the total economy and for the firm itself? Four common procedures may be noted: (1) naive methods, (2) lead-lag series and pressure indexes, (3) opinion polling, and (4) econometrics. In the sections that follow each of these are discussed from the practical standpoint of their predictive value.

NAIVE METHODS

Naive methods of forecasting are typically unsophisticated and unscientific projections based on guesses or on mechanical extrapolations of historical data. As a method of prediction, they may include procedures ranging from simple coin tossing to determine an upward or downward movement to the projection of trends, autocorrelations, and other more seemingly complex mathematical techniques. Typically, they are distinguished from other forecasting methods discussed later in that they are essentially mechanical and are not closely integrated with relevant economic theory and statistical data. Nevertheless, they are widely used by professional forecasters probably because they suggest an air of sophistication and precision to mathematically "naive" executives. Hence, a few of the more common forms should be worth noting.

Factor-Listing Method

One of the earliest forms of forecasting which was common in the 1920's and '30's and is still used by some business firms today may be called the "factor-listing" method. It is worth mentioning first because it presents an interesting point of departure for the discussion of other forecasting techniques, and because it illustrates how "naive" some naive methods can be.

The factor-listing method is a forecasting procedure whereby the analyst simply enumerates the favorable and unfavorable conditions that will affect business activity as he sees them, and then concludes with little or no evaluation or explanation that business will either be good, bad, or the same next year. The method is well illustrated in Table 2–1 which was prepared at the end of 1953 as part of a forecast of the first quarter of 1954.

TABLE 2–1

ROADSIGNS OF BUSINESS, DECEMBER 21, 1953

The major forces affecting the general business outlook for the first quarter of 1954 include:

Favorable	*Unfavorable*
1. *Lower Taxes* On January 1st individual and corporate taxes are to be reduced by $4 billions (annual rate).	1. *Farm Income Off* Farm income is likely to be 10% lower in the first quarter than in the same period this year.
2. *Construction Up* Contracts awarded indicate that construction expenditures in the first quarter are likely to total about $7.5 billions and set a new record high for that quarter.	2. *Inventory Investment Down* Spending on inventory accumulation is likely to decrease by $2 to $3 billions in the first quarter.
3. *Government Spending High* Total government spending is expected to continue high, with a decrease in federal expenditures to be offset by increased state and local government outlays.	3. *Overtime to Decrease* Manufacturing workers are likely to experience a 2% to 3% decline in the work week.
4. *Big Savings Base* Liquid assets owned by consumers have grown by $10 billions, or 5%, in the past year and now total over $200 billions.	4. *Debt Repayment Absorbing More Purchasing Power* Repayments on installment debt in the first quarter are likely to be $2 to $3 billions greater than in the same period this year.

CONCLUSION: THE FAVORABLE FACTORS APPEAR TO BE AS STRONG AS THE UNFAVORABLE, CONSEQUENTLY, CONSUMER DISPOSABLE INCOME IS EXPECTED TO CONTINUE AT ITS PRESENT RECORD HIGH LEVEL OF $250 BILLIONS.

Source. From a paper by R. J Eggert, "How to Forecast Your Company's Sales," delivered at the American Marketing Association's winter meeting, Washington, D.C., December 29, 1953. Mr. Eggert is Program Planning Manager, Ford Division, Ford Motor Company.

Clearly, the list by itself makes no provision for the quantitative evaluation of each of the factors and their role in influencing business activity; it completely ignores the "weighting" of the true forces that have a bearing on business change. Of course the prediction may turn out to be correct, but correct predictions are sometimes realized merely by chance, and this is what distinguishes a forecasting artist from a forecasting scientist. If 1,000 forecasting artists were to attempt, say, to forecast a rise or fall in production by merely tossing a coin, on condition that those who forecast correctly were to forecast again, here is what could happen if at each round the most probable outcome were realized: On the first toss about 500 out of the original 1,000 would have guessed correctly; on the second toss about 250 out of the 500 would be right; on the third toss, about 125; on the fourth toss, about 62 will have called the correct turn; on the fifth, about 31; on the sixth approximately 16; on the seventh, about 8; on the eighth, about 4; on the ninth 2; and on the tenth, 1. This 1 would then hold the record of having called correctly ten turns out of ten—a most remarkable record for a naive method! Obviously, one's "record" as a forecasting artist is not a sole criterion of success, for the most elementary principles of probability dictate that eventually the artist must overplay his luck. If forecasting is to be a science, it must be based on the fundamental assumption that small causes of change, since they are too numerous to measure, will cancel each other out and leave the major causes of change to determine the business trend. And when, as often happens, the small causes fail to offset each other, or develop into unexpected major causes, the forecast will be in error. The forecasting scientist, therefore, can at best hope to be right most (more than half) of the time, as determined by the extensiveness and reliability of the data and his own analytical skill.

Continuity and Trend Models

Proceeding from the purely subjective factor-listing method to more objective naive forecasting procedures, two basic techniques known as continuity and trend projections may be noted first.

1. *Continuity models* are the simplest type of naive procedures used in forecasting. Sometimes referred to as persistence models, they consist of using the last observed variable as a prediction of the future. The underlying assumption, therefore, is that there will be a continuous development of the variable in question. In certain areas and for certain types of short-run forecasting where situations are relatively stable or slowly changing, this method of prediction yields unusually good results. In forecasting weather for certain areas of the country, for example, this technique affords highly reliable answers with a minimum of expense. Insurance companies also employ this method in their construction of life tables, on the assumption that death rates, though not constant, change very slowly. The degree of uncertainty for this type of prediction depends on the stability of the circumstances inherent in the relevant environment.

2. *Trend projections,* as a forecasting method, assume that the recent rate of change of the variable will continue in the future. On this basis expectations are established by projecting past trends, such as least-squares regressions, into the future. This is perhaps the most common method of forecasting used by business firms, not because it is necessarily more accurate than others, but because economic series typically exhibit a per-

sistent and characteristic rate of growth which appears best approximated by a mathematical trend. Accordingly, prediction models of this kind have been used in population forecasting and in stock market forecasting. Companies often project sales several years into the future by this procedure. In basing predictions on trends of past relationships, the trend may be a simple unweighted line, or it may be weighted by attaching greatest importance to the most recent period and successively lesser degrees of importance to periods in the more distant past. Expectational error or variance must also be established, which will be a minimum when the trend is linear.

Trend models have been employed both successfully and unsuccessfully in the past. Forecasts based on 1929, 1933, and 1937 by the method of trend projection were, for companies that employed this method, disastrous. Yet the method continues in wide use, and for a simple reason. Economic time series do, for the most part, show a persistent tendency to move in the same direction for a period of time because of their inherent cumulative characteristics. Therefore, a forecaster using the method of trend projection will be right more times than he will be wrong, and, in fact, he will be right in *every forecast except* those at the *turning points.* Thus, suppose a series rises for 28 months, runs about steady for 2 months, and then declines for 20 months—in all, a total of 50 months. A forecaster using the method of trend projection will forecast correctly the month-to-month direction of change at least forty-eight out of fifty times, which is a score of 96 per cent. And if he tossed a coin for the two uncertain months and guessed one out of two correctly, he could raise his score to 98 per cent—a remarkable record for a naive method! Yet, counting the percentage of correct forecasts appears to be the standard manner of evaluating a forecaster's performance.[1]

Evidently, it is in the prediction of the turning points rather than in the mere projection of trends that the challenge to forecasting really manifests itself. Only when the turning points can be detected in advance can management proceed to alter its plans with respect to sales effort, production scheduling, credit requirements, and the like. Otherwise, the mere projection of trends implies a forecast of continuance and no essential change in policy, and hence the coordinative function of management may become subsidiary to the supervisory one.

Mean and Modal, or Cyclic Models

When the trend is removed from an annual series of economic data, the residual structure exhibits certain fluctuating characteristics that have been described by economists as business cycles. For many years attempts have been made to discover or to prove that a law of oscillation exists in such series, and in some instances the search for periodicity has resulted in outstanding success with respect to prewar series.[2] World War II, however, produced important changes in the structural variables of the economy and altered significantly the phase relationships between time series that had previously exhibited oscillatory characteristics. As a result, cyclic models as a prediction method, though still used by professional forecasters in many business firms, have lost some of their earlier significance. Nevertheless, they contain certain points of theoretical interest which are useful for later discussions in this and subsequent chapters, and hence a brief description of their nature is worth including.

---

[1] C. F. Roos, "Survey of Economic Forecasting Techniques," *Econometrica* (October, 1955), p. 364. See also F. Newbury, *Business Forecasting*, W. Wright, *Forecasting for Profit*, and W. Hoadley, Jr., *Determining the Business Outlook.* Some simple criteria for evaluating a forecast are presented in the concluding sections of this and the following chapter.

In the construction of mean and modal, or cyclic, models, the value of the variable is observed over a long period and, if the variable repeatedly fluctuates over time, the mean of its values is the predicted outcome for the future period. The degree of uncertainty, therefore, depends mainly on the fluctuations about the average. In using this system for prediction, it is necessary to know the expectational error or the difference (deviation) between expected and realized outcomes. These errors may represent the deviation for a single year or the average of deviations over a period of years. This type of model gives a mean expected outcome with a minimum of error when compared to other models, provided that the observations are independent or homogeneous, randomly distributed, and repeated in the "universe." Attempts to predict future occurrences by mean or cyclic models of this nature have been made in the field of astronomy, such as in forecasting comet cycles and sunspots, and in agriculture for predicting yields, weather variations, and insect damage. Some attempts have also been made to forecast stock prices by this method (though unsuccessfully). And the German philosopher Oswald Spengler, in his book *The Decline of the West* (1918), even claimed that every culture passes through a life cycle similar to that of humans, thereby forming what he thought was a basis for predicting the course of civilization.

When the variables being measured are not normally distributed, the modal or most frequent (probable) past outcome may be taken as an expectation of the future instead of the mean outcome. The modal method is thus an alternate way of attacking the forecasting problem, but it gives different expectation values from the mean method when the distribution is skewed or J-shaped. If the distribution is skewed, the average error over a period of years will be less under the mean system than under the mode method; however, a greater number of years in which expectations have a zero error will be realized if the mode system is used. Perhaps this is better understood when it is clear that: (1) the mean outcome may never be realized in a single year, and (2) the modal outcome has a greater probability of occurring than the mean outcome when the data are plotted as a probability distribution (as in Figure 1–2) because the mode is always located under the peak of the curve while the mean is pulled in the direction of skewness. Hence, for a business firm with limited capital that cannot afford the chance of "repeated trials," it would perhaps be better to formulate plans based on the modal outcome rather than the mean if the distribution of expectations is skewed. If the distribution is symmetrical, either method may be used since the mean and mode are identical.

---

[2] See H. T. Davis, *The Analyses of Economic Time Series*, and C. D. Long, *Building Cycles and the Theory of Investment*. For a simplified presentation, see E. R. Dewey and E. F. Dakin, *Cycles, The Science of Prediction*.

Conclusion

In addition to the expectation models outlined above, there are numerous other naive or mechanical systems employed by business firms with varying degrees of mathematical abstraction. One of these, for example, is the method of autocorrelation, where the series is projected by means of a correlation of the series with itself at different points in time, and an "optimum" prediction is then obtained. Regardless of the degree of mathematical abstraction, however, the ultimate success of a forecasting method depends on its ability to include, rather than conceal, the important complex interrelationships that exist in economic series. The method of time-series projections, which is the most widely employed forecasting technique in industry, may serve as an illustration.

Time-Series Projections. At the present time, the use of time-series analysis in forecasting cycles commonly employs what is known as the "residual method." The calculation techniques are described in all elementary textbooks on economic and business statistics and need not be illustrated here. A few words as to the nature and assumptions of the method are of value, however, since the procedure in general plays such a dominant role in the forecasting activities of business firms.[3]

The original data (O) of the series is regarded as being composed of four elements: a secular trend (T), a seasonal variation (S), a cyclical movement (C), and an irregular variation (I). The most common practice is to assume that these elements are bound together in a *multiplicative* structure, so that the relationship is expressed by the formula $O = TSCI$. However, it is also possible to assume that they are *additive* in which case $O = T + S + C + I$, or that there are both multiplicative and additive relationships such as $O = TS + CI$ or perhaps $O = T + SCI$. Various theoretical possibilities may exist, but in most practical problems encountered the multiplicative structure is assumed. In any event, the problem for purposes of forecasting is to isolate and measure each of these four factors, by separating out of the total behavior O, the gradual long-term change T, the regular oscillations S occurring within a year, and the regular oscillations C occurring over several years, each measured independently of the others.

---

[3] There are other approaches used in measuring business cycles besides the residual method, such as the "fixed regularity" method described in Dewey and Daken, *op. cit.;* the "standard cyclical pattern" method which involves an averaging of link-relative changes in important series, developed in E. Frickey's *Economic Fluctuations;* and the generalizations from specific cycles in individual industries and processes as done by the National Bureau in W. C. Mitchell, *Measuring Business Cycles.* The residual method is the one commonly used in industry, however, and hence is the only one discussed here.

The problem of assumed relationships between the series is relatively minor when compared to the other measurement problems that arise.

In explaining the cyclical mechanism, whether for the total economy or for a particular firm, there is a controversy over whether the methods of analysis are really valid. Analysts have shown that apparent cycles can result in a series not because a cycle actually exists, but because of the way in which the data are processed. For example, the use of a moving average may induce an oscillation in a resulting series even if a real cycle is nonexistent, or in general, the summing or averaging of successive values of a random series will result in cyclical behavior by the very act itself (i.e., the "Slutsky-Yule effect"). For these reasons, the conventional method of residual analysis used by most business firms in separating cyclical and random components of time series, is by no means a universally accepted procedure and, as a matter of fact, has been strongly questioned by analysts for many years.

More recently, the separation of trend and random forces in a time series has also been questioned.[4] The assumption in the residual method is that the long-term and short-term movements are due to separate causal influences, so that appropriate mathematical tools may be applied accordingly. But recent studies of economic series reveal that perhaps the trend in a series is not separable from the short-term movements, and that both may perhaps be generated by a common set of forces. Where series of data are observed at fairly close intervals, the random changes from one term to the next may be large enough to outweigh substantially any systematic (causal) effect which may be present, so that the data appear to behave almost like a "wandering series." In such instances it is difficult to distinguish by statistical methods between a genuine wandering series and one wherein the systematic element is weak. Hence, if the series really is wandering, any movements which appear to be systematic, such as trends or cycles, would be illusory, and trend fitting would be highly hazardous.[5]

[4] See, for example, M. G. Kendall, "The Analysis of Economic Time Series," *Journal of the Royal Statistical Society*, Vol. 66, Series A, 1953.

[5] *Ibid.* Kendall points out that an analysis of stock exchange movements revealed little serial correlation (i.e., association between successive values) within series and little lag correlation between series. Therefore, as long as individual stocks behave in the same manner as the average for similar stocks, predicting movements on the exchange a week in advance without extraneous information is impossible. (These statistical concepts are treated further both in a later section and in the next chapter.)

In conclusion, it may be stated that the traditional methods of processing time-series data—methods that are in extensive use by many business firms[6]—are far from adequate, despite their wide acceptance by many professional business forecasters employed in industry. Existing evidence indicates, and further research will probably bear out, that the traditional methods may provide at best a mere description of a set of data and at worst an enormous waste of time, effort, and expense in arriving at what amounts to nothing more than an illusory explanation of systematic movements.[7] Methods exist, however, for meeting some of the shortcomings of traditional time-series procedures, and an illustration proposed by the Cowles Foundation for Research in Economics is presented later in the section on "Econometrics."

LEAD-LAG SERIES AND PRESSURE INDEXES

Two forecasting techniques that are widely used, particularly as a supplement to other forecasting methods, are lead-lag correlations and pressure indexes. Both are relatively simple devices which, when correctly employed, can serve as a useful guide for prediction purposes.

Lead-Lag Series

In the history of forecasting, no method has been given more attention than the lead-lag approach. If a series or index could be discovered that showed leads of, say, six months with substantial regularity, it would indicate successfully the turns in economic activity. Such an indicator would end the quest for a universal "predictor" (and, of course, the need for several thousand professional forecasters and business economists). Forecasters have long sought mechanical methods for predicting the future course of business. Andrew Carnegie used to count the number of factory chimneys belching smoke to tell whether business would rise or decline. The Brookmire Economic Service as early as 1911 utilized successive leads in stock, commodity, and money market series to forecast economic change. And the now defunct Harvard Index Chart, constructed by the Harvard Business School in the 1920's, used a similar set of series to forecast changes in business and finance. In a recent study by Geoffrey Moore and the National Bureau of Economic Research, 801 monthly and quarterly time series or indexes for the United States were examined to determine consistent lead or lag relationships.[8] Moore found the following eleven series to have lead characteristics, where the figures in parentheses represent the average lead in months at business cycle peaks and troughs, respectively:

---

[6] Case studies are presented in National Industrial Conference Board, *Forecasting in Industry* (Studies in Business Policy, No. 77).

[7] Readers with little or no statistical background will find the reference to R. C. Sprowls, in the bibliographical note below, a useful supplement to the above discussion.

[8] G. H. Moore, *Statistical Indicators of Cyclical Revivals and Recessions*, Occasional Paper 31 (1950).

Business failures—liabilities (10.5) (7.5)
New orders for industrial durables (6.9) (4.7)
Residential building contracts (6.2) (4.5)
Common stock prices (6.0) (7.2)
Commercial and industrial building contracts (5.2) (1.7)
Average hours worked per week (3.8) (2.6)
Passenger automobile production (3.0) (4.0)
Wholesale prices, BLS index (2.6) (3.2)
New incorporations (2.5) (3.5)
Pig iron production (2.0) (3.0)
Steel ingot production (1.0) (4.0)

Space does not permit a detailed accounting as to the economic reasons why each of these series tend to lead the business cycle, but a few selected at random may be commented upon briefly. Common stock prices have a long lead because they reflect changes in the demand for funds to finance capital goods, and because investors try to predict future profits of corporations as they buy and sell stock. New orders for industrial goods lead because they reflect businessmen's anticipations of future expansion or contraction. Building construction leads because materials, e.g., concrete, steel, lumber, etc., must first be contracted for before construction can get under way. And basic commodity prices have lead characteristics because manufacturers usually buy raw materials as they receive new orders. But despite these tendencies for certain series to lead, none of them, used singly, can be relied upon for general economic forecasting in a consistent manner; their lead characteristic is not sufficiently consistent over time and with changing economic circumstances. Nor does averaging two or more series seem to offer a practical solution.

Lead-lag techniques are used in various ways by business firms in forecasting their sales. Usually, external correlations are sought between the firm's sales and some outside indicator that is more readily forecast, such as disposable income. Department store sales, for instance, show an average lag of 3.8 and 1.8 months at business cycle peaks and troughs, respectively. Likewise, the demand for children's and baby clothes can usually be predicted, and sometimes also the modal size by age group, by establishing correlations with past population figures and birth rates. Techniques of this and a similar nature are discussed later and in the next chapter in greater detail.

Pressure Indexes

Based largely on the idea that amplitude differences play a significant role in the analysis of business cycles, economists have developed various ratio and difference measures called "pressure indexes" as guides to forecasting. Some examples of such indexes used in economic and business forecasting are the following: (1) Durable goods production fluctuates much more widely than nondurable goods production over the course of a business cycle. Hence the ratio of durable to nondurable goods production is sometimes used as an indicator of cyclical change, the ratio tending to increase in prosperity periods and to decline before a business cycle downturn, although there is no clear-cut evidence of the latter. (2) Purchasing agents, in predicting raw materials prices, frequently use a ratio of raw materials inventories to new orders for finished goods. Also, a somewhat rougher indication is given if production of finished goods rather than new orders is used in the denominator. (3) The difference between the rate of family formation and the rate of housing inventory growth is a pressure indicator of the long-term demand for new housing. In the short run, on the other hand, factors such as disposable income and credit conditions are usually more influential in determining the actual rate of construction. (4) Railroads approximate from six months to a year in advance the demand for new orders for railroad cars from the ratio of carloadings (seasonally adjusted) to cars in serviceable conditions.[9]

These ratio and difference measures, as well as numerous others that can be devised, may not always be helpful in forecasting the magnitude of change. However, they do serve the useful purpose of providing warning signals of impending developments, and frequently an indication as to the future direction of change. When used in conjunction with other forecasting methods, pressure indexes can accomplish much in the way of establishing guideposts for future planning by managers.

OPINION POLLING

The opinion polling technique is a more subjective method of forecasting, amounting largely to a weighted or unweighted averaging of naive forecasts and guesses. The results are arrived at by asking the people who are directly involved as to their future expectations. Various forms of opinion polling are used both in economic and in sales forecasting, and are discussed in the paragraphs below.

Economic Forecasting

The best known opinion polling studies made for forecasting economic activity are: (1) the *Fortune* magazine poll, (2) the McGraw-Hill survey, and (3) the Survey of Consumer Finances. In addition to these, some organizations have made, and some still make, periodic surveys of future business developments. One example is the Dun and Bradstreet surveys, sixteen in all, conducted between the years 1947 and 1951, on business expectations. A number of these surveys resulted in substantially incorrect forecasts for durable and nondurable goods industries, both with respect to trends and turning points. Another example is the quarterly and annual survey of planned business investment made by the Commerce Department and SEC since 1945. The data, based on about 2,500 respondents, provide a tool for short-term projections both in dollar and in real terms, because (*a*) business investment decisions involve commitments in advance, and (*b*) the factors that may cause deviations from plans for individual firms frequently tend to offset one another in the aggregate. Despite these possible advantages, however, sizeable errors in forecasts have been made in past years (especially in 1947 and 1948, for instance) and even the trend has occasionally been incorrect (e.g., 1950) even though the demand for capital goods was already rising sharply when the survey was made.

---

[9] See C. F. Roos and V. VonSzeliski, "The Determination of Interest Rates," *Journal of Political Economy*, (1942), pp. 501–35, and R. Robinson, "Forecasting Interest Rates," *Journal of Business* (January, 1954), pp. 87–100, for other examples of pressure indexes.

The *Fortune* pole is essentially a large, nationwide sample of top executives in medium and large firms. A mail questionnaire is used with responses ranging usually between four and five thousand. Although the percentage of nonresponse is relatively large, experience seems to indicate that among those who respond the rate of refusal on specific questions is low, often only 1 to 3 per cent. The *Fortune* polls have often resulted in correct forecasts as to the direction of movement, but the magnitudes of change have erred seriously. In a statistical sense, this can be explained by the fact that if several hundred respondents indicate the correct moves, and the expectations of the others are close to zero, the directions shown by the survey would be correct, but the magnitudes of change would be in error. Thus, the 1947 *Fortune* miss was probably due to the bias of a weak stock market and the overly pessimistic outlook of some Commerce Department economists. Together, these factors were sufficient to outweigh the correct forecasts made by the few.[10]

The McGraw-Hill survey deals with expenditure plans for plant and producers' durable equipment, i.e., capital-consuming plans. It covers less than 500 companies, but these account for about 60 per cent of the investment of the important capital-consuming industries. The "record" of these surveys has agreed rather well with actual expenditures, except for a few scattered years where the errors could be accounted for by an unexpected reduction in personal income taxes which stimulated demand (1948) or a failure to anticipate the Korean War (1951). Capital expenditure plans, since they are so dependent on changes in the structural environment of the economy, could not be expected to remain the same under such unusual circumstances. Other than these, however, the McGraw-Hill surveys have provided a basically sound analysis of capital expenditure plans. They cover much the same ground as the government survey mentioned above, but are available earlier (published in *Business Week* magazine) and are widely used for forecasting purposes.

The Survey Research Center of the University of Michigan has been sponsored by the Board of Governors of the Federal Reserve System to prepare surveys of consumer finances and buying plans. The surveys, whose sample sizes range from 2,900 to 3,600 cases, are designed to: (1) evaluate recent developments among consumers, (2) provide data for testing hypotheses about economic behavior, i.e., functional relationships between variables, and (3) determine expectancies for consumer purchases of automobiles, houses, and major appliances. A single survey provides a cross section of data while consecutive surveys yield time series of such data. The results of these surveys have not been promising as a guide to forecasting consumer demand. In general, consumers appear to be relatively insensitive to small changes in prices and incomes. Threats of shortages usually bring stronger reactions than do small price changes, at least with respect to durable goods purchases. There have been times, e.g., 1946–49, when the direction of year-to-year changes did not correspond with actual changes for durable goods purchases. And in reinterviews conducted less than a year after the initial interview, substantial percentages of respondents (sometimes well over 75 per cent) did not recall having ever declared their expectancy to buy a particular durable. Apparently, the average consumer is not a very rational planner; his decisions are affected by a wide array of economic and emotional complexities which he cannot unravel and use as a basis for future buying plans. He often guesses at, rather than plans for, his future purchases. If the changes in basic demand determinants (such as price, income, consumer stocks, etc.) are small, then in a probability sense the guessers will have an average expectancy of zero, and the buying trend will be determined by the few who do plan on the basis of their confident expectations.

---

[10] See Roos, "Survey," p. 384.

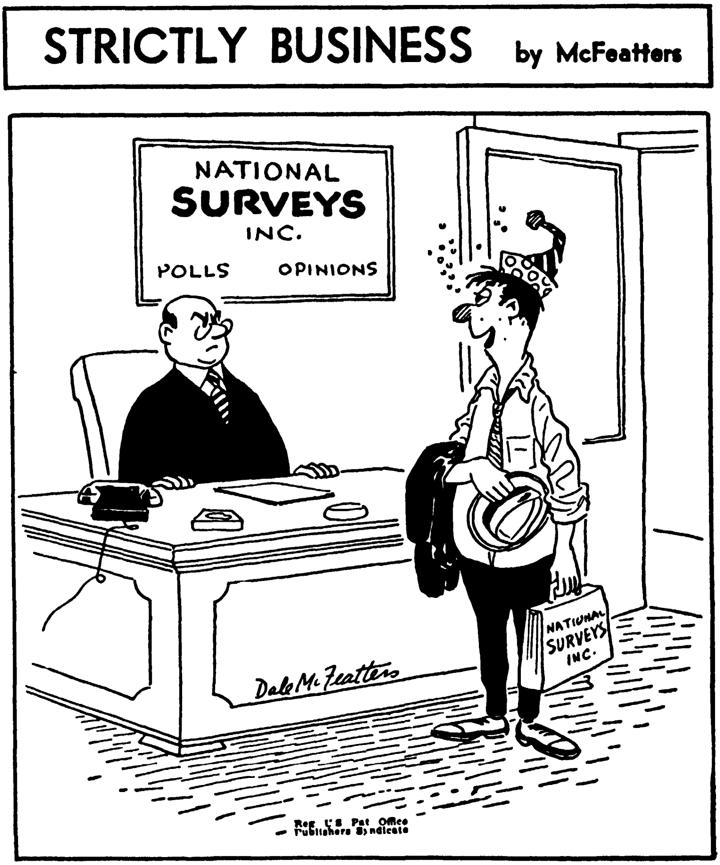
Sales Forecasting

Several variations of opinion polling methods are used by companies in forecasting sales.

1. *Executive polling* is frequently employed, whereby the views of top management are combined and (subjectively) averaged. The assumption in the use of this approach is that there is safety in numbers, in that the combined judgment of the group is better than the forecast of any single member. Hence the executives sit as a jury and pass judgement on the sales outlook for the coming year. Generally represented on the jury are those with a divergency of opinions—the sales, production, finance, purchasing, and administrative divisions. In those companies where forecasts of probable events are derived after a sifting and analysis of market reports, sales data, and economic forecasts, the executive polling approach has been fairly successful. Without such careful evaluations, however, the method can easily degenerate to the level of a guessing game yielding nothing more than sloppy and unfounded predictions. Companies employing the executive polling approach frequently combine it with statistical measures of trends and cycles (i.e., naive methods, discussed

above) as a further tool, by raising or lowering the statistical forecast according to their subjective judgement.

2. *Sales force polling* is another variation whereby a composite outlook is constructed on the basis of information derived from those closest to the market. The sales forecast may be built up from salesmen's estimates in cooperation with branch or regional managers, or by going directly to jobbers, distributors, and major customers in order to discover their needs. The advantage claimed for the method is that it utilizes the first-hand, specialized knowledge of those nearest to the market, and thereby gives salesmen greater confidence in their quotas developed from forecasts. On the other hand, salesmen may be quite unaware of structural changes taking place in their markets, and hence incapable of shaping their forecasts to account for those future changes. Also, there is the realistic fact that sound forecasting requires more time and effort than most


STRICTLY BUSINESS by McFeatters
NATIONAL SURVEYS INC.
POLLS OPINIONS
NATIONAL SURVEYS INC.
Dale McFeatters
Reg. U.S. Pat. Office Publishers Syndicate

"I found 138 persons watching TV, 29 listening to radio and 3 having a party!"

salesmen can ordinarily devote, and the result may be an off-the-cuff guess rather than a prediction based on careful reasoning. Accordingly, firms using this method have usually set up a system of "checks and balances" whereby salesmen's estimates are compiled, checked, adjusted, and revised periodically in the light of past experience and future expectations.

3. *Consumer intentions surveys* are still another version of the opinion polling method applied to sales forecasting. The Ford Division of the Ford Motor Company, for instance, makes sample surveys of automobile buying intentions which it then projects to a national level by weighting the estimate with the average purchase rate (about 1.4 per family) and an index of predicted incomes. Similar techniques are used by other firms in forecasting the sale of appliances, furniture, and other durable goods.

New Products

Various forms of opinion polling have been widely used to forecast the demand for new, as compared to already established, products. In fact, where new products are concerned, the public opinion approach is quite common because little or no past data exist. There are at least two ways in which the problem can be attacked, separately or in combination.

One method is to take a sample of expected purchasers and determine their relevant demand characteristics, e.g., price, income, need, etc. For consumer products, demonstrations are often helpful in estimating these characteristics (e.g., early pressure cookers). For an industrial good, a sales engineer equipped with drawings and specifications can sometimes get an indication of need from an interview with a potential company buyer. In any case, inferences based on probability can be made about the entire universe (market) once the parameters (central tendency, confidence limits, etc.) of the sample distribution are established.

A second method is to survey specialized dealers who are presumably close to the consumer and know his needs and alternatives. In the early days of the electric blanket, for instance, manufacturers polled both regular blanket dealers and electrical appliance retailers to obtain an estimate of demand in potential markets. Polling dealers, however, can also result in wide error margins, especially where new products are concerned; dealers usually have a tendency to be either overly enthusiastic or unduly pessimistic, as revealed by the actual results later on, and supplementary forecasting methods are usually advisable before a final estimate is reached.

ECONOMETRICS

Forecasting is a science of prediction. Economic forecasting is the science of predicting economic change. It is because change occurs that the need for forecasting arises, for in a stationary economy where the future is merely an extension of the present, forecasting would be unnecessary: what happened yesterday would happen today and would also happen tomorrow. But change does occur, economic activity is not static, and therefore managers must make predictions if forward planning is to be successfully accomplished.

Based on the idea that changes in economic activity can be explained by a set of relationships between economic variables, there has grown up a branch of applied science known as *econometrics.* Breaking the word into two parts, "econo" and "metrics," it is evident that its subject matter must deal with the science of *Economic measurement.* And this is precisely what econometrics does: it is a method of explaining past economic activity, and of predicting future economic activity, by deriving mathematical equations that will express the most probable interrelationship between a set of economic variables. The economic variables may include, for example, disposable income, money stock, inventories, government revenues and expenditure, foreign trade, and so on. By combining the relevant variables, each a separate series covering a past period of time, into what seems to be the best mathematical arrangement, econometricians proceed to predict the future course of one or more of these variables on the basis of the established relationships. The "best mathematical arrangement" is thus a model which takes the form of an equation or system of equations that seems best to describe the past set of relationships according to economic theory and statistical analysis. The model, in other words, is a simplified abstraction of a real situation, expressed in equation form, and employed as a prediction system that will yield numerical results. To the extent that economic theorems and relationships can be verified by subjecting historical data to statistical analysis, then, at least in principle, econometrics as a system of measurement stands midway between the theorizing of the "ivory tower" (pure) economists and the extreme nontheoretical empiricism of the radical institutionalists.

Statistical Aspects

As should be evident from the previous chapter, economic theory deals with the science of choice between alternatives, and its method is to construct simplified models of economic reality on the basis of which certain laws describing regularities in economic behavior are derived. When these models are quantitatively constructed, they are known as econometric models. Such models have been constructed for the total economy with the objective of predicting the future levels of employment, income, and other general economic variables. They may be labeled *macro*econometric models in order to distinguish them from *micro*econometric models. The latter are constructed for a particular firm or industry rather than for the total economy (i.e., macroscopic as compared to microscopic) and are designed to predict the particular firm's or industry's sales, production, costs, and related economic variables, the future courses of which are desired by executives as a guide for improved decision making. It is the micro-type model, of course, that will occupy the core of our attention in later chapters. However, both types of models involve certain statistical implications, some knowledge of which is essential if correct interpretations of the results are to be made. Let us examine these implications briefly.

Statistical analysis is evidently an indispensable part of econometrics. Modern statistics, as most readers will recall even from an elementary course in the subject, is based upon the concept of probability. As already indicated in the previous chapter, there is even some disagreement among leading writers on the subject as to exactly what is meant by probability. One school of thought (e.g., Keynes and Jeffreys) views it as a system of logical propositions rather than as an empirical process; another (e.g., von Mises and Teller) believes that it deals with the limit that is approached by the relative frequency of an event as the number of trials increases indefinitely. For the practical problems typically encountered in econometrics, it is the second concept employing an empirical approach—probability being the limit of a relative frequency—that is the more relevant one, and hence the one that is assumed in some of our later discussions.

Recognizing probability as the most fundamental concept upon which statistical analysis is based, we may outline three aspects of the subject with which the reader is always assumed to be familiar when he interprets the results of econometric investigations, as we shall be doing in a few of the later chapters.

1. *Statistical Inference.* This is the science of drawing conclusions about a population on the basis of information derived from a sample. All economic relationships, particularly from an econometric standpoint, are regarded as samples from an unknown infinite population of all possible economic relationships, and econometricians employ statistical methods in order to arrive at numerical estimates.

2. *Randomness.* Basic to the entire science of statistical inference is the notion of a randomly determined variable, or as statisticians call it, a "stochastic" variable. This is a variable that can assume a number of values with given probabilities as, for instance, when stated in the previous chapter that, with continuous rolls of a perfect die, it can be established with certainty that the probability (i.e., limit of the relative frequency) is $1/6$, or 0.17. The interpretation of randomness may be thought of, therefore, as meaning that any given value of the variable is *independent:* it is not in any way determined by its past values nor does it in any way determine its future values. The importance of this will be discussed in later chapters where, for example, in forecasting a time series of company sales, the values may not be entirely independent or, in other words, not randomly distributed in the manner of a stochastic variable, with the result that "serial correlation" is said to exist.

3. *Point Estimation.* This is a very important part of statistical inference and hence of econometrics. In point estimation, the problem is to derive or predict a single figure as an estimate of the unknown quantity.

Two methods commonly used for this purpose are the method of least squares and the method of maximum likelihood. The former, that of least squares, is more frequently employed by practicing forecasters, and is the familiar procedure learned in elementary statistics. It is based on the principle that the sum of the deviations squared when taken from the mean is a minimum (hence the term "least squares") and it chooses the estimate which minimizes the sum of the squared deviations from the chosen value. The method of maximum likelihood, on the other hand, chooses the value which makes the probability of occurrence of the estimate a maximum. Both methods, however, often give the same estimates, and some results of each method will be illustrated in later chapters where empirical studies are discussed.

Thus, with probability as the underlying concept, these three features—statistical inference, randomness, and point estimation—combine to form the methods, materials, assumptions, and objectives of modern statistics, and hence are the dominant characteristics of econometric analysis. For econometrics is concerned with the numerical evaluation and statistical verification of economic laws or relationships. In a world of uncertainty, econometrics explicitly introduces random variations and leads therefore to probabilistic conclusions. And even though the factual data which serve as the raw material of econometrics are rarely precise, and thereby lead to errors of observation, such errors, quite realistically, are often regarded as random, and methods exist for including them in the general probability scheme of econometric analysis. In short, the objective of econometrics is to predict the most likely course of future economic events on the basis of relationships revealed in the past, by a sensible method of extrapolation that is not naive or mechanical.

Correlation

Modern statistical analysis, as we have just seen, incorporates statistical inference, randomness, and point estimation among its most important features, and these combine to form the skeleton of econometrics. But econometrics with this skeleton alone is like the Pauper without the Prince. The skeleton must be clothed with flesh to make it live, and the tool or *modus operandi* for accomplishing this is the powerful statistical device known as "correlation analysis." The general concept and technique of correlation will be illustrated by an ingenious graphic method in the following chapter, so that it may be comprehended by the nonmathematical reader whose only equipment need be a knowledge of arithmetic and an eye for curves. But on the basis of what should already be familiar from an elementary course in statistics, some simple but essential concepts may be presented at this time, since correlation is perhaps the most important statistical tool used by econometricians.

The purpose of correlation analysis is to arrive at a mathematical equation, called an "estimating equation," "predicting equation," or "regression equation," which best discloses the nature of the relationship that exists between a dependent variable and one or more independent variables. Variations in the dependent variable, such as the sale of automobiles, are to be predicted on the basis of variations in one or more independent variables, such as income, number of families, replacement rates, and so on. The independent variables chosen are the factors that are believed to be the controlling ones on the basis of economic theory.[11] The ultimate predicting equation that emerges is derived on the basis of statistical or correlation analysis. Where only one independent variable is involved, the statistical analysis is known as *simple correlation;* where two or more variables are involved, the analysis is called *multiple correlation.* Conceptually, the relationships may be illustrated by the following example.

If we let Y denote the sales of a product and X its price, then variations in Y will depend upon variations in X, so that the relationship could be written conceptually as $Y = f(X)$. This is read "Y is a function of X," or "sales are a function of price," and states merely that a dependent or functional relationship exists between the two variables. Of course, any other letters could be used in place of those chosen without affecting the meaning. In this equation, since only one independent variable is involved, the relationship is one of simple correlation. Similarly, if further analysis revealed that not only price, but other factors, such as advertising and income, also had an important influence on sales, the functional relationship could then be written $Y = f(X, W, Z)$ where W and Z now refer to advertising and income, respectively. The equation would now read, "Y is a function of X, of W, and of Z," or "sales are dependent upon price, upon advertising, and upon income." And, of course, just as previously, other letters could also be used if desired to signify the same fact that a relationship exists. In any event, since more than one independent variable is involved, the relationship is that of multiple correlation.

Now the purpose of correlation analysis, whether simple or multiple, is to derive the actual equation of relationship among the variables rather than the conceptual one stated above. This amounts to saying that the separate independent variables must be weighted in some way, according to their proper importance in their effect upon the dependent variable. Thus, in a case of simple correlation, the relationship between the dependent and independent variable might appear to be best represented by a straight line. The desired equation would then be of the form $Y = a + bX$, which is the equation for a straight line, with Y the dependent variable and X the independent variable. The letters a and b for the particular relationship, however, are not known: they are the weights (the constants or parameters) whose values are to be determined, and it is the purpose of correlation analysis to arrive at their most probable values.

---

[11] It should be noted that academic economic theory is only one among alternative sources for developing hypotheses. Though it forms a good starting point, all econometric analyses need not be built upon it.

When derived, they might, for example, be found to be $a = 8$ and $b = 3$. Then the predicting equation would be $Y = 8 + 3X$, and by forecasting the value of X for any given period in the future, the corresponding value of Y can be predicted by simply substituting for X in the formula. (E.g., at $X = 5$, $Y = 23$; at $X = 10$, $Y = 38$; etc.)

This, very briefly, is the conceptual nature of correlation analysis and its function in econometrics. These and many related concepts will be treated more fully in the following chapter so that the profound implications of the technique may be readily seen and appreciated.

Types of Models

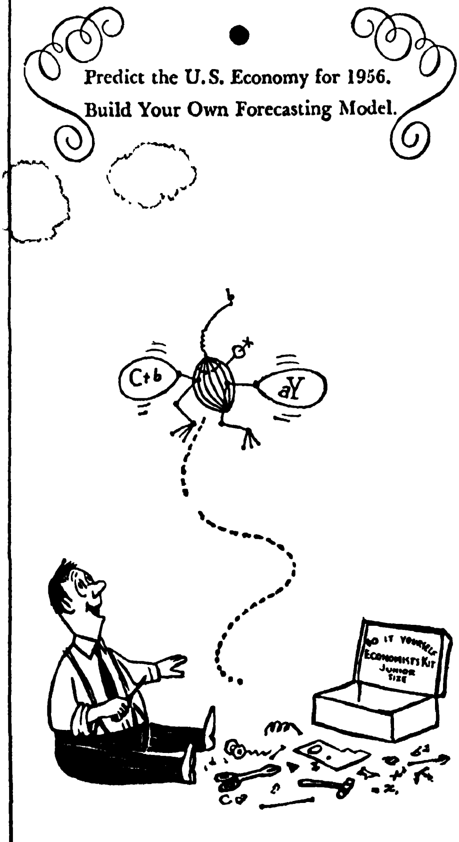
The use of econometric models—mathematical equations expressing economic relationships—for purposes of forecasting, involves the process of deriving quantitative relationships between essential variables in order to predict the outcome of an economic event. The economic event may be the level of employment, income, etc., for the total economy, or it may be sales, production, costs, etc., for a particular firm or industry. Econometrics has thus been employed as a forecasting method both at the macroeconomic (aggregate economy) level and at the microeconomic (individual firm) level. Although the applications of the latter will occupy most of our attention throughout this book, some comments about both are in order at this time.

It may be recalled from a knowledge of elementary statistics that, in correlation analysis, a single equation is derived which is then used for making predictions. This equation, as noted above, expresses the relationship between a dependent and one or more independent variables. However, in recent years, some econometricians operating at the macroeconomic level have centered their attention around the use, not of single equations for prediction purposes, but of *simultaneous equations* or systems of equations instead. Their argument is that this is the correct method to be used when estimating the structural relationships between many variables which themselves may be intercorrelated. Thus, in a single-equation model, it is assumed that the relationship is: (1) one way in nature so that variations in the independent variables cause variations in the dependent variable, but not vice versa, and (2) that the independent variables are themselves not significantly intercorrelated so that their separate variations may be measured without confounding the results. In most macroeconomic forecasts, this group argues, these conditions are not met: the separate variables are significantly intercorrelated so that the single-equation model yields incorrect results; therefore, a system-of-equations model is preferred where each equation in the model expresses a distinct and essential set of relationships.

At the microeconomic level, on the other hand, most of the econometric analyses have been of the single-equation type. Aside from the additional time and expense considerations of constructing a simultaneous-equation system, there are at least two other reasons that account for the use of one-equation models.

1. The relationship between the dependent and independent variables is usually one way in nature. In forecasting the sale of a consumer good, for example, an independent variable might typically be disposable income or some other general economic variable. In this case, it is reason-

The Junior Econometrician's Work Kit.


Predict the U.S. Economy for 1956.
Build Your Own Forecasting Model.
C+b
aY
DO IT YOURSELF
ECONOMISTS KIT
JUNIOR SIZE

DIRECTIONS:

1. **Make up a theory.** You might theorize, for instance, that (1) next year's consumption will depend on next year's national income; (2) next year's investment will depend on this year's profits; (3) tax receipts will depend on *future* Gross National Product. (4) GNP is the sum of consumption, investment, and government expenditures. (5) National income equals GNP minus taxes.

2. **Use symbols for words.** Call consumption, C; national income, Y; investment, I; preceding year's profits, P_{-1}; tax receipts, T; Gross National Product, G; government expenditures, E.

3. **Translate your theories into mathematical equations:**

(1) $C = aY + b$
(2) $I = cP_{-1} + d$
(3) $T = eG$
(4) $G = C + I + E$
(5) $Y = G - T$

This is your forecasting model. The small letters, a, b, c, d, e, are the constants that make things come out even. For instance, if horses (H) have four legs (L), then $L = aH$; or $L = 4H$. This can be important in the blacksmith business.

4. **Calculate the constants.** Look up past years' statistics on consumption, income, and so on. From these find values for a, b, c, d, and e that make your equation come out fairly correct.

5. **Now you're ready to forecast.** Start by forecasting investment from this year's profits. Look up the current rate of corporate profits — it's around \$42-billion. The model won't tell what federal, state, and local governments will spend next year — that's politics. But we can estimate it from present budget information — it looks like around \$75-billion.

6. **Put all available figures into your model.** (We've put in the constants for you.)

(1) $C = .7Y + 40$
(2) $I = .9 \times 42 + 20$
(3) $T = .2G$
(4) $G = C + I + 75$
(5) $Y = G - T$

7. **Solve the equations.** You want values of C, I, T, G, Y. Hints: Do them in this order — (2), (1), (4), (3), (5). In solving (1), remember that I and E are both part of G, $Y = G - T$, and $T = .2G$.

8. **Results.** (See if yours are the same.) For 1956, consumption will be \$260.0-billion; investment, \$57.8-billion; GNP, \$392.8-billion; tax receipts, \$78.6-billion; national income, \$314.2-billion. These results are guaranteed — provided that the theories on which they're based are valid.

Reprinted with permission from *Business Week,* September 24, 1955 able to suppose that variations in income will cause variations in sales, but variations in the latter in turn will not cause significant variations in the former.

2. The variables can usually be combined in such a way that the intercorrelations are insignificant. Thus, two highly correlated variables such as the price of food and the cost of living can be combined as a simple or weighted ratio or average and this new synthetic index can be used as an independent variable in a regression equation. Quite stable results can thus be secured as will be illustrated in Chapters 5 through 7. At present, this is sufficient to explain the nature of econometrics as a method of forecasting until the following chapter where, after a discussion of graphic multiple correlation, the single-equation and multiple-equation problem, as well as others, will be discussed more fully.

Conclusion

In concluding this section, a comment as to the significance of econometric models in forecasting is in order. In one such macroeconometric model,[12] an attempt was made to predict disposable national income from fifteen endogeneous variables (jointly dependent variables within the process) and fourteen exogeneous variables (predetermined variables or those outside the process). Fifteen equations entered into the estimative procedure and the model was compared with two naive models of the continuity type. Naive model I stated that next year's value would equal this year's value plus a random disturbance; naive model II stated that next year's value would equal this year's value plus the change between the two years plus a random disturbance. As it turned out, the econometric model failed for certain conditions to provide a better prediction for the year 1948 than did the less costly and much simpler naive models.

Does this mean that econometric methods should be abandoned? The answer is probably no. Where our theoretical understanding and statistical data are good, econometrics can illuminate the darker areas and enhance our ability to predict. Since forecasts usually must be quantitative, econometrics, crude as it may be, is best adapted for obtaining quantitative results. In any forecast, there are always certain strong forces that can come into play and modify trends. The econometrician knows this and he constantly measures the intensity of new forces as they appear. And it is this constancy of observation that enables him to identify the turning points in advance. At the macroeconomic level, the present goal of econometric methods is not to develop a comprehensive model, but to continue research on the quantitative measurement of economic change. But whether at the macroeconomic or microeconomic level, it should be kept in mind that no mathematical procedure, regardless of how complex, can ever become a substitute for genuine sophistication. This is particularly true of naive methods, perhaps the most important use of which is that they provide a benchmark—a null hypotheses—against which the more sophisticated forecasting methods can be compared, and will continue to be used by business firms until better prediction tools are devised.

---

[12] See Carl Christ, "A Test of an Econometric Model for the United States, 1921–47," in *Conference on Business Cycles* (National Bureau of Economic Research, 1951).

CHOICE OF A FORECASTING METHOD

Now that various alternative methods of forecasting available to management have been surveyed, what criteria may be established as guides to the selection of an appropriate method? The standards listed in the paragraphs below are qualitative in nature and hence they complement the quantitative criteria to be outlined in the concluding section of the following chapter. Since no particular forecasting method is best suited for all companies, a decision to choose one method rather than another should eventually depend on management's own evaluation of the relevant choice indicators. Some of the key indicators based on subjective considerations are as follows.

1. *Comprehensibility.* Management must have some understanding, and appreciation of the method used if the forecasts are to be employed as a basis for planning. Forecasts are of little value if no confidence is placed in the procedure employed to obtain them. This has been a serious factor limiting the use of mathematical methods in many firms, and perhaps a major reason for the wide use of naive and subjective methods in practical work.

2. *Accuracy.* The most difficult part of forecasting is to predict the turn rather than just project the trend. The latter merely reflects a momentum that carries business along in the same general direction as the over-all business cycle, and a forecast of continuance implies no change in plans. As the turn is approached, however, new decisions are necessary and new plans must be formulated to account for changes in raw and finished goods inventory, hiring and training of personnel, capital expenditure schedules, and sales effort. Accuracy, therefore, should be measured not only in terms of the percentage of correct forecasts (since scores close to 100 per cent can be achieved in this way as shown earlier), but more so by the ability to predict the turning points.

3. *Timeliness.* The forecasting method chosen should utilize the most recent data available, and should be flexible enough to incorporate necessary new data to meet changing conditions. Management is thus provided with current or near-current information as a basis for planning.

4. *Usability.* The method chosen should yield forecasts that are in units and product groupings readily usable within the company. In production scheduling, for instance, the forecast should be in appropriate physical units or readily convertible into such units. For the finance department, the forecast should be available in current dollars, deflated dol-

lars, or in any other form for which it is needed, and similarly for other departments and divisions of the company.

5. *Economy.* Finally, the method chosen should be within the staff and expense limitations of the firm, though these are often difficult to establish. The limits, however, will depend largely on the values management places on sound forecasting as against the costs of such forecasts. Often, a company's past experience with formal methods of forecasting is a major consideration in choosing the method most suitable for its own purposes.

In the final analysis, a successful forecasting operation requires not only the most suitable method, but also interdepartmental cooperation and a management that is confident in the results. Without the necessary cooperation, an adequate forecast is impossible; and without management's confidence in the forecast, the prediction, no matter how sophisticated, is useless.

BIBLIOGRAPHICAL NOTE

A discussion of opinion polling methods in forecasting is available in G. Katona, *Contributions of Survey Methods to Economics*, especially Chapter II. The significance of the Survey Research Center techniques is treated in the *Federal Reserve Bulletin* (July, 1950), and in the appendixes for the June, 1947 and June, 1949 issues. I. Schweiger, "The Contribution of Consumer Anticipations in Forecasting Consumer Demand," *Short Term Economic Forecasting*, is another useful source, as well as F. Modigliani and O. H. Sauerlander, "Economic Expectations and Plans of Firms in Relation to Forecasting," *Short Term Economic Forecasting* (National Bureau, 1955), and I. Friend and J. Bronfenbrenner, "Business Programs and Their Realization," *Survey of Current Business* (December, 1950). Some recent contributions along other lines are the following: J. H. Lorie, "Forecasting the Demand for Consumer Durable Goods," *Journal of Business* (January, 1954), as well as that particular journal issue as a whole which is devoted entirely to the subject of forecasting; a general and typically scholarly article by C. F. Roos, "Survey of Economic Forecasting Techniques," *Econometrica* (October, 1955), which brings up to date many of his thoughts expressed earlier in his well-known little book *Charting the Course of Your Business*, and also the same author's *General Outlook for the American Economy, 1954–74.* On econometrics, two standard works are G. Tintner, *Econometrics*, and L. Klein, *Textbook of Econometrics*. Readers with a year or more of statistical analysis will find E. G. Bennion, "The Cowles Commission's 'Simultaneous Equation Approach:' A Simplified Explanation," *Review of Economics and Statistics* (February, 1952), a useful source providing a brief exposition of modern econometric thinking. An amplification of the article is presented by J. Meyer and H. Miller, Jr., "Some Comments on the 'Simultaneous-Equations' Approach," *Review of Economic Statistics* (February, 1954). A general article on the subject is C. Christ, "Aggregate Econometric Models," *American Economic Review* (June, 1956).

Less technical and more survey-type discussions of forecasting appear in a number of readily available sources, including E. Bratt, *Business Cycles and Forecasting,* 4th ed.; M. Colberg, Bradford, and Alt, *Business Economics,* rev. ed.; J. Dean, *Managerial Economics;* and J. Howard, *Marketing Management.* Case studies of company practices are found in the National Industrial Conference Board's *Forecasting in Industry* (Studies in Business Policy No. 77). Also, in a light vein and humorously written is a special report article, "Business Forecasting," *Business Week* (September 24, 1955). Finally, those interested in the elementary statistical implications will find a clear discussion in various chapters of Bross, *Design for Decision,* and in an excellent introductory text by R. C. Sprowls, *Elementary Statistics,* chaps. 11-13 dealing with time series.

QUESTIONS

1. In your own words, define briefly the various forecasting methods discussed in this chapter.
2. What is wrong, if anything, with the "factor-listing" approach to economic forecasting?
3. (*a*) Distinguish between a continuity or persistence type of naive model, and a trend model. (*b*) Which is more common in economic and business forecasting? Why?
4. Trend models will usually yield correct forecasts, at least as to direction of change, more often than not. If the distinction between a forecasting artist and a forecasting scientist is that the latter is correct more than half the time, the use of trend models would seem warranted. Yet we have been critical of their use. Why?
5. "After all, in the final analysis, the best forecasting method is obviously the one that yields the highest percentage of correct predictions." Comment.
6. The "residual method" is probably the most common approach employed in industry to forecast cyclical fluctuations. Comment on the use of this method and its significance.
7. State two ways in which lead-lag series are sometimes used as a guide to sales forecasting by business firms.
8. (*a*) How useful are lead-lag techniques for forecasting the total economy, i.e., macroeconomic forecasting? (*b*) What about forecasting for a particular firm, i.e., microeconomic forecasting?
9. What is a "pressure index"?
10. For an industry, the ratio of inventory stocks to sales is a pressure index. What managerial decision might this ratio portend if it were relatively high?
11. (*a*) In general, which has been more successful as a forecaster: the McGraw-Hill survey of capital expenditure plans, or the Survey of Consumer Finances? (*b*) Explain why, from an economic standpoint.
12. In what ways are opinion polling methods applied by firms for their own sales forecasting?
13. (*a*) Why are opinion polling methods often more suitable for predicting the sales of new rather than established products? (*b*) What survey approaches are usually employed.

14. Of the various forecasting methods discussed in this chapter, what is unique about econometrics?
15. The "trinity" of econometrics is (*a*) statistical inference, (*b*) randomness, and (*c*) point estimation. Explain. Where does probability theory fit in?
16. (*a*) Distinguish between simple correlation and multiple correlation. (*b*) What is the role of correlation analysis in econometrics?
17. Write in conceptual form using your own choice of symbols: "The sale of automobiles depends upon income, prices, and the number of households."
18. (*a*) When might a simultaneous-equation model be preferable to a single-equation model? Why? (*b*) Which type of model is most common, especially in business forecasting?
19. What important functions are performed by naive-model forecasting methods?
20. In the final analysis, what are the important choice criteria to consider in selecting a forecasting method?

Chapter 3

ECONOMIC MEASUREMENT

It should be evident by now that the over-all problem confronting business managers is one of *adjusting to uncertainty*. This was first indicated in Chapter 1 where we outlined the state of uncertainty as an environmental characteristic of the real world, and whose existence makes executive decision making and forward planning a necessity. It was emphasized further in Chapter 2 where we pointed out that the first step in adjusting to uncertainty is the adoption of an appropriate forecasting method that would facilitate the process of decision making and forward planning. Now, in Chapter 3, we turn our attention to the concluding step in the adjustment sequence: the methods and techniques available to managers for establishing quantitative estimates of economic relationships, or, in short, the science of economic measurement.

What is the purpose of discussing the techniques of measurement in a book devoted to the economics of business management? The answer is obvious. Management wants and needs to have essential relationships and predictions expressed in quantitative terms if it is to formulate plans involving the hiring of x number of workers, the scheduling of y units of production, or the marketing of z units of output. In other words, plans must generally be reduced to numerical terms if they are to guide the decision makers whose function it is to steer the firm's course into the future. The chief measurement techniques commonly used to derive the needed estimates are discussed briefly in the sections that follow. The manner in which the topics are treated, as in the rest of the book, is essentially literary rather than mathematical, so that the fundamental concepts can be readily grasped by those possessing nothing more than an eye for curves and perhaps an ability to handle some "advanced arithmetic."

MEASUREMENT METHODS

Several basic procedures are available to analysts engaged in business research. The choice as to which method to use depends on the particular problem at hand, the quantity and quality of available data, the time allotted to the research, and the budget allowance. Any one of these factors can swing the balance in favor of an alternate, though less acceptable, procedure. In actual business research, academic perfection must often be compromised with practical business considerations. This is one of the facts of life that usually causes much grief and frustration for the inexperienced researcher fresh out of graduate school.

There are many approaches to economic measurement, but usually they are modifications or variations of four basic methods: (1) sample surveys, (2) controlled experiments, (3) accounting and engineering methods, and (4) correlation analysis. Each of these can often be used separately or in combination with others, and are discussed below with this understanding in mind.

Sample Surveys

The use of sample surveys, though employed in various fields, is probably the most common technique in market research, where the purpose usually is to obtain information of a demand nature from a "representative" group of the population under study, e.g., housewives or retailers. The information may reveal data as to buying intentions, behavior, motivations, or other vital facts upon which to base rational decisions. If the survey is well constructed, it can also be a useful means for intelligent guessing as to the influence of particular demand determinants. For example, a study made for a chain of jewelry stores helped reveal the extent to which price and income variations were influential in affecting sales. Since income differences in potential markets had to be established, they were estimated from the latest county records on assessed valuations by geographic areas. On this basis, along with some supplementary information, management was able to make rough estimates of price and income elasticities for each of its stores in each city. Price-quantity relations were also established by income groups from what consumers said they had previously paid, as well as what they would be willing to pay, for various kinds of jewelry.

Consumer surveys as a type of sample survey can sometimes be of help in deciding on a price range for a new product. In 1954 when Armour & Co. was preparing to market its breakfast beef as a substitute for bacon, the question as to how the product should be priced relative to premium bacon had to be decided. The information obtained from interviews provided a first approximation as to the appropriate price spread to maintain during the product's introductory phase.

Surveys of buyers' intentions have been made, of which the Federal Reserve Board's annual consumer-finance study is the best known. These surveys are designed to reveal the intentions of consumers as to their future purchase of durable goods. Although the survey approach when employed for purposes such as these is not too reliable an indicator of potential demand (as discussed earlier in the previous chapter), it may be helpful when used in conjunction with more refined analytical methods. A recent development that may prove valuable in this respect is the field of "motivation research," in which psychological depth interviews are employed in order to discover the factors that motivate people to buy. But the success of this approach will ultimately depend on the ability of psychologists to develop new and better experimental methods and measurement techniques.

Sampling Considerations. From the measurement standpoint, a major problem confronting the decision maker is the sampling procedure employed in the sample-survey method. Traditional sampling procedures have used what might be called the "sample estimation" approach, where the purpose of the sample has been to estimate the population (universe) characteristics, e.g., a consumer survey seeking to determine the relative popularity of various brands of cigarettes. To accomplish this, the size of the sample is predetermined according to established principles of sampling theory. In the final analysis, what is wanted is a representative sample. This means that the sample must be large enough so that opposing errors will cancel out (i.e., the problem of sample size), and that the sample must represent adequately the relevant segments of the population in the necessary proportions (i.e., the problem of sample design). Meeting both of these problems and overcoming them may often prove to be costly and time consuming. Accordingly, a radically new type of sampling procedure called *sequential analysis* has been developed, as explained in Chapter 1, whereby substantial savings in money and time can be achieved relative to the conventional fixed-size sample. Some additional comments about the method may be of interest at this time.

Sequential analysis involves a method of sampling whereby the same accuracy can be achieved as with the conventional random sample, but the sample size can often be reduced by more than half and sometimes by even more than 60 per cent. Essentially, the difference between conventional sampling and sequential analysis is this: in conventional sampling, no final (or terminal) decision is made until all the sampling is completed; in sequential analysis, the size of the sample is not predetermined but depends instead on the values of the observations. The analysis is conducted in stages and a decision must be made at each stage as to whether to go ahead or make a terminal decision. In practice, after each sample observation or group of observations is secured, the result obtained from the accumulated observations is compared with a previously calculated set of acceptance and rejection numbers, and on this basis a decision is made as to whether to stop or to continue sampling. On the average—that is, the average for repeated sampling—the sequential method requires a smaller sample size to reach a given decision than would be required by ordinary simple random sampling. But it sometimes happens that sequential sampling results in a larger sample than simple random sampling with equal accuracy. Generally speaking, however, sequential methods are preferable to nonsequential types of analysis when: (1) the data are available serially (e.g., telephone interviews) so that the results of the observation are known in a minimum of time, and (2) the cost of the data (in terms of time, labor, etc.) is approximately proportional to the amount of the data.

Despite these limiting factors, statisticians seem to believe that the application of sequential analysis to problems of sample estimation is hardly more than a matter of time.[1] Eventually, it will be possible to estimate population values from a sample within a predetermined range and with a given probability without first having to know the population variance. But even today, where sequential analysis can be employed, as in industrial inspection problems, substantial savings in time and money can be realized relative to conventional sampling methods.

Controlled Experiments

While questionnaire interviews may be helpful in obtaining product preferences and customer intentions, the fact remains that there is often a difference between what people say they will do and what they actually do. In determining demand the chief concern is not with intention, but rather with actual purchases of a given commodity. Based on the conviction that present actions are a better guide to the prediction of future action than are intentions or beliefs, researchers have developed controlled observational methods of analysis sometimes called *controlled experiments* or *experimental design.* From a statistical standpoint, two basic steps are involved in this type of analysis: (1) the variables or determinants to be manipulated must be separated from each other and from all residual variations, and (2) an adequate basis must be established for computing error. Several ingenious methods for accomplishing these ends have been devised, of which one of the most common in business research is the *Latin square.* A brief illustration of its use is presented in the following paragraphs with reference to a marketing problem.

In a study made for the Market Basket Corporation, a supermarket chain in western New York State, the problem was to discover the effects of different pricing, displaying, and packaging practices on the demand for McIntosh apples. For example, a question to be answered was whether apples should be sold in 4-, 5-, 6-, or 8-pound Polythene (transparent-type) bags. The problem was further complicated by the fact that day-to-day variations in sales of apples had to be accounted for: grocery sales as a whole are larger on weekends, and it had to be determined whether the larger sale of apples on Saturday as compared to Tuesday, for instance, was attributable to (function of) a different-size package or was merely a reflection of generally increased grocery sales. A second factor for consideration was the particular store. Apple sales differ between stores, and these differences had to be accounted for in order to attribute the variations in apple sales to the size of the bag. The experimental design that was employed is shown below.

---

[1] See R. Ferber, *Statistical Techniques in Market Research,* p. 183; J. H. Lorie and H. V. Roberts, *Basic Methods of Marketing Research,* pp. 155–57; and M. J. Moroney, *Facts From Figures,* chap. 12.

EXPERIMENTAL DESIGN, APPLE STUDY

First Part of Week					Second Part of Week				
Day	Store				Day	Store			
	1	2	3	4		1	2	3	4
Mon.	B	C	D	A	Fri. A.M.	B	A	C	D
Tues.	A	B	C	D	Fri. P.M.	C	D	B	A
Wed.	D	A	B	C	Sat. A.M.	A	B	D	C
Thurs.	C	D	A	B	Sat. P.M.	D	C	A	B

The stores are numbered 1, 2, 3 and 4 (since four stores were used) and the treatments or bag sizes are lettered A, B, C, and D. Each treatment is given once in each store on each day during the week, so that on any one day, all four treatments are used, and over the whole time period (one week), each store is exposed to all four treatments. As can be seen, the design consists of a double grouping of variables to be treated, with the arrangement of the treatments among the stores and time intervals distributed *at random.* By rotating the treatments to be tested, both row and column differences (i.e., differences in day or week and store as demand determinants) are eliminated, and the change in apple sales due to different bag sizes measured.

This design measures the effects of three demand determinants, packaging, day of week, and particular store. Other designs can be constructed, however, to measure the effects of two factors, or even of four or more factors. In the above study, for example, the design was extended to include the effects of promotional methods (window displays and advertisements), bulk sales methods (price in two-pound units and four-pound units), and quality variations (bruising and color). Whereas traditional methods of experimenting require that all factors but one be held constant and then vary that one to measure its effects, the experimental-design approach permits and, in fact, encourages multiple variations of factors and, by the method of covariance analysis (correlation and variance analysis) measures the separate and combined effects of the factors. Depending on the arrangement, designs, if properly constructed, can yield a maximum amount of information from a minimum number of observations.[2]

---

[2] The above design was a 4 by 4 type, i.e., four rows and four columns. Latin squares, however, may in practice be as large as 8 by 8, but beyond this they become too inconvenient to handle.

In the apple problem discussed above, it was also decided to determine whether

Similar experiments have been conducted by other companies for such products as costume jewelry, tires, and appliances. For each problem there were different variables to be measured, such as price spreads, brand differences, color, etc., and their effects on sales. As a research method, controlled experiments of the type discussed are still relatively new and untried in the commercial field. This is due partly to the difficulty of planning the study in such a way that the relevant variables can be segregated and measured. Being able to determine which variables are most important in a particular problem so that the experiment can be properly planned is a major part of the job. Sample surveys and good critical judgment in the preliminary stages of the research can save much time, work, and expense later on.

In demand problems, controlled experiments may be dangerous since they can cause unfavorable reactions on the part of consumers. And in oligopoly situations, they may create retaliatory actions by competitors, particularly if the study is conducted on a large scale in major markets. But despite these obstacles, experimental design promises to become a major method of commercial research in the years to come.

Accounting and Engineering Methods

For certain types of problems, accounting and/or engineering methods of measurement are often used. Usually, these approaches to measurement emphasize the collection, classification, and interpretation of data on the basis of inspection and experience in the case of accounting methods, and the nature of physical relationships between such factors as man-hours, capacity, and pounds of output where the engineering method is used. Accounting and engineering approaches are most commonly employed in establishing production and cost estimates, and in profit planning by the use of break-even charts (discussed in Chapter 4), but they have also been used with various degrees of success in other areas of company management and planning. They frequently have the advantage of being less expensive than other measurement methods, often because they require less variations in the underlying data. Hence they are more useful as preliminary estimates or as supplements to more comprehensive analyses, rather than as an end in themselves. Because of their essentially noneconomic nature, they will be given relatively less attention in this book except to point out the areas in which they are commonly used, and how they may sometimes be integrated with other measurement procedures as a tool for more effective control by manage-

---

a change from one treatment to another had any effect within the stores. Accordingly, a special "double change-over" design was constructed, consisting of two orthogonal Latin-square designs in which the sequence of treatments was reversed in the two squares. This procedure eliminated store and time variance and at the same time permitted the measurement of carry-over effects.

ment. As a brief illustration, however, the following accounting-engineering approach to the solution of a problem is fairly typical.

When the demand for certain types of products is estimated, especially industrial goods, the savings in costs that can accrue to the buyer (due to greater efficiency, for example) are frequently a chief factor in determining the demand for the good. Demand schedules may thus be constructed by translating the buyer's benefits into actual dollar-and-cents cost savings. This approach can also serve as a useful guide to the seller as to whether the product should be sold outright or made available on a rental basis. The steps followed by the accountant or engineer in estimating the cost savings that may accrue to a potential buyer of an industrial good, such as a machine, might be as follows: (1) A common basis for measuring the comparative worth of the two machines (i.e., the new and the old) is chosen. The type of base might be a unit of time (such as a week, month, or year), output per man-hour, or any other relevant common denominator. The ultimate purpose is to arrive at a figure showing the total expense per unit of the base chosen, such as total expense per average week of operation for each machine, or cost per machine per hundred hours of operation, or cost per machine per thousand units of output, etc. (2) All of the costs entering into the operation of each machine according to the base chosen are computed. For example, if the base is a year, it is necessary to calculate for each machine on an annual basis such cost factors as depreciation, insurance, power, light, heat, repairs, supplies, attachments and accessories, and interest (since the machine represents an investment to the buyer).

From these estimates of the cost savings involved, a rough schedule is constructed relating expected sales to prices. If the good is to be sold to more than one buyer, i.e., a number of firms in one or more industries, the problem is somewhat more difficult, since cost savings estimates are probably impossible to obtain for all prospective buyers. An alternative, therefore, is to select a representative sample of (potential) company purchasers, perform the same basic calculations as above, and then construct an expected sales schedule for the total market on the basis of the sample. In establishing the sample criteria for "representativeness," factors such as output rates, wage rates, efficiency, or similar characteristics should be considered.

This method, as will be seen later, is somewhat different from the approach suggested by economic theory, where expected earnings on capital is a crucial determinant of equipment purchases in this kind of demand analysis. That is, to the economist, existing equipment held by the buyer should be viewed as a historical cost; decisions to buy new equipment should depend on the relative worth of alternatives as measured by the discounted value of prospective earnings on the investment choices open to the firm. This is discussed further in Chapters 10 and 11 below, dealing with capital management.

Correlation Analysis

One of the most powerful tools of economic measurement and the one most widely used by econometricians is correlation analysis. The purpose of correlation analysis, as stated in the previous chapter, is to formulate a mathematical equation which best expresses the relationship between essential variables so that prediction may be facilitated. The data required for the analysis are typically obtained from historical records or from controlled experiments, although the former source has been the more prevalent one.

Correlation analysis is particularly well suited to agricultural products as compared to manufactured goods, which is a chief reason why most of the pioneering contributions to econometrics came from econometricians working in the field of agricultural economics. Farm products are for the most part homogeneous, and they are sold in competitive markets where price fluctuations are wide and relatively easy to measure. For manufactured goods, on the other hand, the products are largely heterogeneous and their prices remain rigid for relatively longer periods of time. Accordingly, in making a demand study, for instance, it is often necessary to measure the influence of other factors along with price, such as income, advertising, and other demand determinants, in order to arrive at a meaningful demand function that can be useful for prediction. Correlation, however, has been successfully employed and is gaining increasing importance in manufacturing, particularly in the areas of demand, production, and cost measurement. The results of a number of outstanding studies in these fields will be presented and discussed later in Chapters 5, 6, and 7.

The standard mathematical methods of correlation, especially multiple correlation, are adequately described in many textbooks on statistics. However, a nonmathematical method based on graphic procedures has been developed. This graphic method, which is far easier to comprehend than the mathematical method, can be performed in about one third to one half the time, and if done carefully can yield results that are good approximations of those involving the use of mathematical equations. In view of these considerations and their obvious value in practical work, plus the fact that the graphic method of multiple correlation is not described in most of the standard works on statistics, the remaining sections of this chapter are devoted to an explanation and discussion of the concept. An understanding of the method will prove useful both as a valuable measurement technique and as a basis for comprehending several other topics on measurement to be discussed in subsequent chapters.

GRAPHIC MULTIPLE CORRELATION

In any prediction problem it is desirable to find the relationship between two or more variables. It may be useful to measure, for instance, the relation between a company's sales on the one hand, and the Federal Reserve Index of Industrial Production, disposable income, and advertising outlay on the other. In this section a simple technique for measuring such relations will be explained, based on the use of nothing more than tabular and graphic methods. The same basic procedure is applicable to the measurement of relationships in a variety of areas in business economics, as will be seen subsequently once the method is understood.

For illustrative purposes, the example to be developed is a sales forecasting problem based on three series of hypothetical data. These are a demand series, a price series, and an income series. The dependent variable, demand, will be designated as Y and is measured in number of units sold; the two independent variables are X_1 and X_2 for price and income, respectively. A time or observation period of ten years will be covered, for which the relevant data are shown in the first four columns of Table 3-1. The double line in the table separating these first four columns from the remainder of the table signifies that the information presented at the right, i.e., columns 5 through 9, must be derived from the given informa-

TABLE 3-1

TABULAR FORM FOR GRAPHIC MULTIPLE CORRELATION

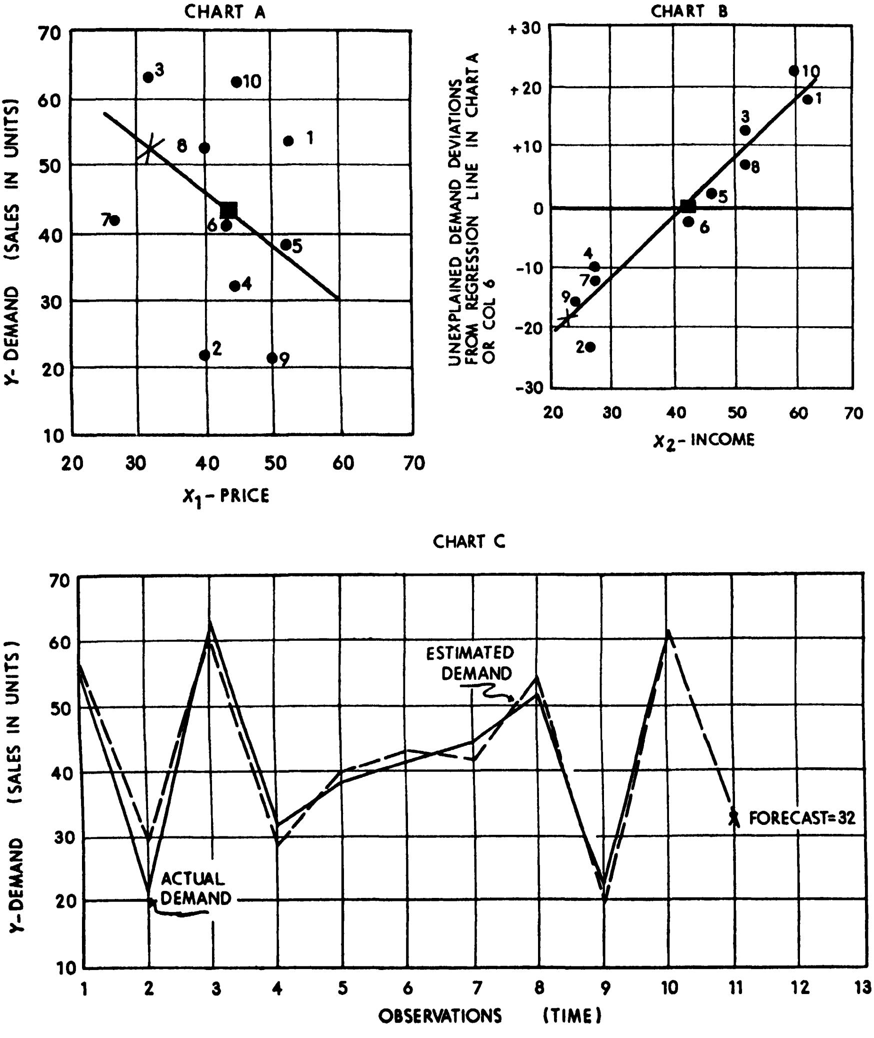
(1)	(2) Y	(3) X_1	(4) X_2	(5)	(6)	(7)	(8)	(9)
Observation Year	Demand (Thousands of Units)	Price (Dollars per Unit)	Income per Family (Hundreds of Dollars)	Expected Demand As Estimated from Regression of Y on X_1 (Chart A) (Regression Estimates As Read off Vertical Axis)	Actual Demand Minus Expected Demand (Unexplained Deviations) (Chart A) (Vertical Deviation of Dots from Regression Line) (2) − (5)	Expected Additional Change in Demand Due to Income (Chart B) (Regression Estimates As Read off Vertical Axis)	Demand As Estimated from Both Price and Income (Chart C) (5) + (7)	Actual Demand Minus Expected Demand (Chart B) (Vertical Deviations of Dots from Regression Line) (2) − (8)
1	54	52	62	36.0	18.0	20	(56.0)	−2.0
2	22	40	26	45.0	−23.0	−16	(29.0)	−7.0
3	64	32	52	51.0	13.0	10	(61.0)	3.0
4	32	44	28	42.0	−10.0	−14	(28.0)	4.0
5	38	52	46	35.5	2.5	4	(39.5)	−1.5
6	42	42	42	43.5	−1.5	0	(43.5)	−1.5
7	44	26	28	56.0	−12.0	−14	(42.0)	2.0
8	52	40	52	45.0	7.0	10	(55.0)	−3.0
9	22	50	24	37.5	−15.5	−18	(19.5)	2.5
10	62	44	60	42.0	20.0	18	(60.0)	2.0
Total	432	422	420		−1.5			
Mean	43.2 $\bar{Y}$	42.2 $\bar{X}_1$	42.0 $\bar{X}_2$					
11 (Forecast)		(32)	(22)	(51)		(−19)	(32)	

tion in columns 1 through 4. Essentially, the problem is to discover a method for forecasting the dependent variable, demand (or Y), on the basis of the independent variables, price (X_1) and income (X_2). Thus we are assuming that a functional relationship exists which may be written in general form for conceptual purposes as $Y = f(X_1, X_2)$, or using the initials for demand, price, and income, as $D = f(P, I)$. Both formulations are equivalent. They state that a functional or dependent relationship exists between the demand for this particular product on the one hand, and its price per unit and income levels on the other.

In the mathematical method of correlation, the object is to arrive at

FIGURE 3-1

GRAPHIC MULTIPLE CORRELATION



the *equation* which best explains the nature of this relationship. In later chapters dealing with the econometric measurement of demand, production, and cost relationships, numerous examples of such equations (functions) will be illustrated. In the graphic method of correlation, the object is to arrive at the *curve* which explains the relationship, as will become apparent in the following paragraphs. Also, it should be emphasized again that although the discussion involves a demand problem, the general method to be explained is applicable to all types of forecasting where the problem is to predict the value of a dependent variable on the basis of changes in one or more independent variables.

Finally, a word as to the method of presentation before beginning the actual analysis. The discussion will be pitched at two levels: (1) at an elementary level for the benefit of readers who have little or no knowledge of multiple correlation analysis (such as those who have never gone beyond an elementary course in business statistics), and (2) at a slightly higher level for those who may already have a moderate reading (but not necessarily working) knowledge of the subject. For the latter group some comments labeled "remark" will be interspersed at various points; the beginner with little or no knowledge of correlation analysis can safely skip over these comments in a first reading. However, in rereading and studying the method, these remarks should be included since they form an integral part of the analysis as a whole, and will provide a much stronger foundation for comprehending the results of various econometric analyses in later chapters.

Arithmetic Means (Table 3–1)

The first step in the procedure is to compute the totals for the three series and then their arithmetic means, as was done in columns 2, 3, and 4. The symbols $\bar{Y}$, $\bar{X}_1$, and $\bar{X}_2$ represent the respective means thus computed. These symbols are read "Y-bar," "X-bar sub-one," and "X-bar sub-two," respectively, the "bar" being commonly used in statistics to represent a mean.

Scatter Diagram (Chart A)

The problem is to explain or account for the changes in demand in terms of price and income changes. Accordingly, a scatter diagram is plotted of the dependent variable Y against one of the independent variables, such as X_1. These are the dots in Chart A of Figure 3–1. The order of choice of the independent variables will not affect the analysis, although it is customary to arrange them in decreasing order of importance, such as X_1, X_2, etc., according to their effect on the dependent variable.

Remark. The order of choice can be facilitated by plotting preliminary scatter diagrams between the dependent and each of the independent variables.

The independent variable which appeared to have the highest correlation with Y would then be designated as X_1, the next highest as X_2, and so on in that order. Sometimes the relationship can be seen directly from the table so that separate preliminary diagrams are unnecessary to establish the order.

Since the dots must later be identified, they should be labeled by placing next to them the number representing the corresponding observation period from column 1 of the table. The scatter of the dots, as it now stands, represents the simple correlation between Y and X_1. It reveals that there is a tendency for demand to be high when price is low and for demand to be low when price is high.

Regression Line (Chart A)

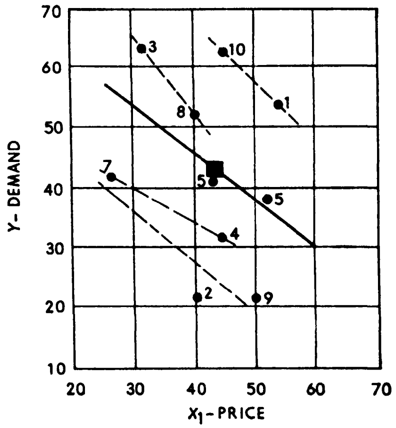
The next step is to pass a freehand line through the dots in such a manner that it seems to represent best the pattern of the scatter. The line may be either curved or straight, depending on what is believed to be the best representation of the pattern. (The implications of this are discussed in the next section.) The meaning of the line will be the same regardless of the type of line, although if a curve is employed there is the further complication of judging the correct type of curvature. In Chart A a straight line was used since there did not appear to be a marked curvilinear relationship. In drawing this line, two guideposts are of some help: (1) If a straight line is used it should be made to pass through the average values of Y and X_1. These were computed earlier as the first step in the analysis, and were recorded near the bottom of columns 2 and 3 of Table 3–1. For Y and X_1, they were 43.2 and 42.2, respectively. These two averages are plotted against each other as a single point and represented on Chart A as a filled-in black square. The line, as can be seen, passes through this point.

Remark. The reason why the line should pass approximately through this point is because of a least-squares principle that all multiple regressions intersect the mean values of the variables. It should be noted, however, that in drawing a curved regression line, the best fit is *not* the one which passes through the means of the two series. Only for straight lines does the means rule apply.

(2) Since this is an average line and cannot pass through all of the observations, a second guidepost is that the line should be drawn so that about half the dots are above and half are below the line. Thus, in Chart A, there are five dots above the line and five below. It should be emphasized that, for this second guidepost particularly, it is quite sufficient to approximate the slope or steepness of the line, because any error made will be corrected in the next chart.

Remark. To facilitate establishing this second guidepost, which really amounts to estimating the slope of the line, analysts customarily construct what is known as "drift lines." These lines are derived as follows: (1) Groups of two or more observations for X_2 that have approximately the same values are chosen. In this problem, as can be seen from Table 3–1, there are four such group observations: 1, 10; 2, 7, 9; 3, 8; and 4, 7. (2) Each of these groups of observations are then connected by a single freehand regression line. Since there are four such groups in this problem, there would be four such regression or drift lines. (Colored pencils or dashed lines should be used to distinguish these drift lines from the over-all regression line.) These drift lines are easily drawn since they usually connect only two observation points in each group; where three or more points occur in a group, they usually fall almost along a straight line, since constant or nearly constant values of X_2 were chosen. As many such drift lines as possible should be drawn. (3) The average slope of all these drift lines combined would then approximate the slope of the over-all regression line. In the accompanying diagram, Figure A, which is a reproduction of Chart A, the drift lines are the dashed lines connecting the four sets of observation points noted above. The solid, over-all regression line should then have a slope about equal to the average slope of the drift lines combined. The use of drift lines represents the graphic counterpart of estimating the values (slopes) of the net regression coefficients in the mathematical method.

FIGURE A



Regression Estimates (Chart A)

As it stands, the regression line in Chart A represents the average or mean relationship between Y and X_1 (demand and price) with X_2 (income) constant.

Remark. It should be stressed again that the accuracy of this first approximation should not be judged by the closeness of fit of the regression line to the scatter of Y on X_1, because the partial regression (Y on X_1 with X_2 constant) will equal the simple regression (Y on X_1) only when the correlation between X_1 and X_2 is zero. This is not usually expected in a sample.

Accordingly, we can estimate the demand that may be *expected* (not necessarily realized) from variations in price with income held constant. Thus, from Chart A, we observe that for year 1, when price was 52, the demand as would be expected from the regression line was 36, and this figure is then recorded in column 5 for the corresponding year. In like manner, the remainder of column 5 is filled out for all the remaining observations, by reading off Chart A the expected demand for each observation as determined from the regression line.

Unexplained Deviations (Chart A)

Since the plotted points in Chart A do not fall closely along the fitted line, it is apparent that a considerable portion of the fluctuation of demand is unexplained by its relationship to price. Numerous other factors in addition to price may be influencing the demand for this commodity. The prices of competing products, income fluctuations, advertising, and changes in buying habits of people may be just some of the factors exerting an influence on demand. In order to seek a further explanation, therefore, it is necessary to determine the amount of variation that is unexplained by the association of demand and price. In year 1, for instance, both Chart A and Table 3-1 show that the expected demand as estimated from the regression line was 36 while the actual demand realized was 54, a difference between actual and estimated of +18. Similarly, for year 2, the actual demand was 22 while the expected demand was 45, a difference of −23. These figures are recorded in column 6 of the table. When read from Chart A, they represent the *vertical* deviations of the dots from the regression line; when derived from the table, which is simpler once column 5 is completed, they are simply column 2 minus column 5.

Regression Line (Chart B)

In order to explain further the variations in demand unaccounted for by price, we turn to the next independent variable, income or X_2. Chart B is now constructed by scaling off on the vertical axis the deviations just recorded in Column 6 of the table, while the horizontal axis ticks off the income series from Column 4. Note that a horizontal line, representing the zero value for the deviations, is drawn through the chart (usually near the center), with the plus deviations above this line and the minus deviations below. (This zero line is the regression line drawn horizontally.) A scatter diagram is again plotted, this time between the Column 6 deviations and the income variable of Column 4, and the resulting dots are again labeled with their corresponding observation number. A freehand regression line is fitted to the dots, as before. Note also that the line passes through the filled-in black square which this time represents the mean of X_2 at the zero line.

Remark. Actually, the total of the demand deviations in Column 6 would have been zero instead of −1.5 as shown (and hence the mean would have been zero), had the regression line in Chart A been a perfect fit. For it is a mathematical characteristic of the mean that the sum of the deviations about it is zero. As it happened, the scatter in Chart B fell fairly closely about the line, but this would not have occurred had the sum of the deviations been much larger than −1.5. With practice the analyst can learn to approximate the first regression closely. Had drift lines been used to begin with (which they were not) in fitting the over-all regression line in Chart A, the scatter in Chart B would have been still smaller. Unfortunately, drift lines tend to clutter up the chart and are often more confusing at first than they are helpful. In view of this, it was felt that the explanation as a whole would gain more in clarity than it lost in accuracy by confining the discussion of drift lines to "remarks." In any case, if the first regression passes through the means of Y and X_1, the second (if linear) will pass approximately through the mean of X_2 at the zero line. Incidently, it is also customary to use drift lines based on constant values of X_1 in fitting the second regression line.

In economic terms, the regression line in Chart B represents the relationship between demand and income after allowance has already been made for the influence of price on demand. That is, it tells us the amount by which demand will be higher or lower than the demand-price regression line of Chart A according to different levels of income. To illustrate, in Chart A with price (X_1) at 40, the demand (Y) may be expected to average about 45 if income (X_2) remains at its average value. But in the year 2 when the price was actually 40, the level of income (X_2) was sufficiently below its average to pull the demand down to 22 as compared to what would have been expected from the partial Y-X_1 relationship. The same idea is shown perhaps more clearly by the table. In year 2, demand was 22 (column 2) as compared to the expected 45 (column 5), because income (column 4) was 26 as compared to its average of 42. In the year 1, on the other hand, the actual demand was higher than the expected demand based on the Y-X_1 relationship, because income was sufficiently above its average value to raise the actual demand to 54 instead of the demand of 36 that would have been expected on the basis of price alone as a demand determinant. Hence the second independent variable may be regarded as a "demand shifter," since the regression line in Chart A may be thought of as shifting up and down by a number of units equal to that indicated by the corresponding level for X_2. These further effects of income on demand are shown in column 7, which is derived by taking the expected regression values off the vertical axis of Chart B.

Forecasting (Chart C)

How can the graphic method of multiple correlation be used in forecasting the dependent variable, demand? From the managerial standpoint, it is the application of the method to prediction and measurement that determines its usefulness for decision making and planning. Hence, an indication of how well the technique can be applied to forecasting provides a test as to the usefulness of the analysis.

As signified by the word itself, the term "correlation" (or "co-relation") refers to the nature of the interdependency that exists between two or more variables. The purpose of correlation analysis is to establish this interdependency in numerical terms, i.e., to quantify the relationship. Once this has been done, the problem of predicting the dependent variable is nothing more than a mechanical application or extension of the basic relationship already established.

In either method of correlation analysis, whether mathematical or graphic, the independent variables must be predicted first, and on the basis of these predictions the dependent variable is then forecast. Thus, correlation analysis does not solve entirely the problem of prediction; it merely facilitates forecasting the dependent variable once the independent variables have already been estimated for a given period in the future, and the nature of the functional relationship is known. The forecast procedure, therefore, may be accomplished as follows.

1. In column 8 of the table, record the demand that would be expected on the basis of the partial Y-X_1 relationship (with X_2 constant) and the demand that would be expected on the basis of the partial Y-X_2 relationship (with X_1 constant), or simply the sum of columns 5 and 7. These figures represent the demand or sales that would be expected on the basis of variations in *both* independent variables, price and income, and are recorded in parentheses to emphasize the fact that they are estimated rather than actual or realized figures.

2. On Chart C, plot the actual sales variations of Y (column 2) as a solid line and the final estimated sales variations (column 8) as a dashed line. These figures are plotted as a time series with sales or demand on the vertical axis and the observation periods or time on the horizontal. The two lines as they now stand provide a visual representation of how well we were able to estimate sales as compared to the actual sales variations that occurred.

3. It remains now to make a prediction of sales, say for year 11, on the basis of the relationship already established. Suppose that for year 11, therefore, it was estimated (predicted)[3] that for the independent variables, X_1 would be 32 and X_2 would be 22. What may sales be expected to be on the basis of these price and income values? The first step is to record the two price and income predictions at the bottom of columns 3 and 4, respectively, on a line with year 11, the forecast year. Note again that the figures are written in parentheses to distinguish them as estimated rather than actual values. We then turn to Chart A and note that when price is 32, the expected demand based on the regression line (shown by an "X" on that line) is 51, and this number is written in parentheses at the bottom of column 5 of the table. This is what demand is expected to be on the basis of the partial Y-X_1 regression. Similarly, turning to Chart B, we note that when income is 22, the expected value based on the regression line (shown by an "X" on that line) is -19, and this figure is then recorded in parentheses at the bottom of column 7 of the table. This is what we would expect sales to be on the basis of the partial Y-X_2 regression. We now have two sales estimates, the first based on price with income con-

---

[3] Prediction of the independent variables is a separate problem area the discussion of which is deferred to sections in this and subsequent chapters.

stant, the other based on income with price constant. To arrive at a forecast for year 11, these two estimates are brought together by recording their total, 32, in parentheses at the bottom of column 8. As with the other numbers in this column, the figure 32 represents the net estimate of demand based on both price and income. This value is then recorded on Chart C for year 11 as an "X," and connected by a dashed line to the previous actual figure. This completes the forecast. In like manner, sales can be predicted for any year in the future, provided of course that price and income can be forecast first for the same time period.

Remark. The final dashed line connects the predicted demand with the last actual demand, because it is the actual that is being forecast. When year 11 is over and actual sales realized, the latter would be connected by a solid line to the previous actual sales, and the forecast sales connected by a dashed line to the previous estimated sales.

It is possible to extend this method to include three or more independent variables in the same manner as the two used in this problem. If an additional variable were included, say X_3, the residuals in column 9 would be obtained by reading off the deviations about the regression line in Chart B, or more simply by taking column 2 minus column 8. These values are then plotted against the new variable X_3 and a new set of estimates obtained. The regression line would then represent the partial relationship between Y and X_3 with X_1 and X_2 constant, and the final sales estimate would reflect the combined effect of the three independent determinants, X_1, X_2, and X_3. Stated conceptually, the general functional form would be $Y = f(X_1, X_2, X_3)$ instead of $Y = f(X_1, X_2)$ as was assumed in the problem discussed in this section. A forecast of sales, therefore, would require that the three independent variables be predicted first. In practical work, however, few problems often require more than perhaps four independent variables, and many can be handled adequately with three or less. Moreover, analysts are not usually concerned with obtaining a perfect fit, i.e., accounting for the entire variation in the dependent variable by the independent factors included. Instead they are quite content to explain a substantial part of the total variation in the dependent variable, say 90 per cent or more, and this can frequently be done by incorporating one, two, or three of the most important independent factors.

Remark 1. Statistically, a high degree of multiple correlation, i.e., a nearly perfect fit, would be of no significance if obtained by too complex an analysis relative to the number of observations available, because there would be too few degrees of freedom. Accordingly, analysts usually employ straight lines or logarithmic curves based on a minimum of about 20 years of data.

Remark 2. It should be noted again that in drawing a curved regression line, the best fit will be obtained with a line that does *not* pass through the means of the two series. As stated in an earlier remark, the means rule applies only for straight lines.

Further Refinements

The method of graphic correlation analysis discussed above is sometimes called "the method of successive approximations" because it provides a means by which the first approximation to the regression lines can be progressively improved. Although the procedure in this problem employed only a first approximation because the results seemed quite adequate for the purpose at hand, the remaining steps for approximating the regression lines more accurately may be sketched briefly.

1. The vertical deviations from the regression line in Chart B are recorded in column 9 of the table. These deviations would then be plotted *about the regression line* in Chart A against X_1. (Colored pencils may be used so as not to confuse this new set of dots with the original ones.) The resulting dots would form a new scatter and the deviation for each observation would be directly above or below the original dot for that observation, since the same X_1 values were retained. A new regression line (the same color as the new dots) would then be drawn through this new scatter so that it passes approximately through the means shown by the original filled-in square, and in such a manner that it appears to reduce the scatter about itself to a minimum. This new regression line would represent an improved approximation to the partial Y-X_1 relationship.
2. The vertical deviations of the new scatter about the new regression line in Chart A would then be plotted about the regression line in Chart B against the same X_2 values, and this time a new regression line on Chart B would again be drawn through the filled-in square. This new regression line on Chart B would represent an improved approximation to the partial Y-X_2 relationship.

The above two steps may be repeated, using the latest deviations about the latest regression approximations, until no further correction seems necessary. In the illustration used here, only a first approximation was made so as to convey the basic concept and technique, and because no further approximation appeared desirable for improving forecasting accuracy as shown visually by Chart C. If the charts are published, only the last set of observations about the final regression line should be shown, since it represents the most refined estimate. The construction of Chart C would be essentially the same as before, except that actual sales would be compared to estimates based on the latest approximations.

Remark 1. Two further points may be noted. (A) It can be proven that each approximation moves closer to the mathematically calculated least-square value at a very rapid rate, the lower the intercorrelation between the independent variables. (B) The new scatter in Chart A indicates the part correlation between Y and X_1. If the intercorrelation between X_1 and X_2 is low, the part correlation will almost equal the partial correlation which shows the degree of relationship between Y and X_1 with X_2 constant.[4]

Remark 2. The use of systems of equations, or simultaneous equations, may be justified where intercorrelation is significant. This is discussed further in this and in subsequent chapters.

Estimates of Determination. As an indication of the extent to which the dependent variable is explained by the independent variables for the years included in the analysis, a measure called the *coefficient of multiple determination* (symbolized R^2) can be computed. The formula is:

$$R^2 = 1 - \frac{\Sigma(d^2)}{\Sigma(Y^2) - (\bar{Y})^2 N} \geqq 0 \leqq 1 ,$$

where

Σ = the Greek capital letter sigma, meaning "the sum of,"
d = the deviation of any point from the final regression in the last chart (Chart B) or the difference between the actual and estimated value for Y (column 9),
$\bar{Y}$ = the mean of the Y values,
N = the number of observation in the analysis.

This coefficient gives the percentage of variation in the dependent variable which is explained by the independent variables in the analysis, and hence yields a solution between 0 and 100 per cent (expressed in decimal form). The closer the coefficient is to 1, the closer the dashed line will be to the solid line in Chart C. The multiple coefficient of determination is also of extreme practical value to management from the standpoint of investment in research. As will be pointed out in later chapters, adding further variables to the analysis will increase the coefficient (the marginal returns to precision) at a decreasing rate, while it increases the costs of computation in terms of time and expense (the marginal cost of research) at an increasing rate. Hence it is not the increased accuracy alone which is a deciding factor as to whether more variables should be included in the analysis, but the *increment in gains* realized from the greater accuracy balanced against the *increment in costs* involved. Evidence of this will be show in Chapters 5 through 7. Incidentally, it may also be mentioned that the square root of the multiple coefficient of determination is the coefficient of multiple correlation, R, which is an abstract measure and not as useful for practical purposes as is the squared form.

A second measure is called the *coefficient of partial determination* (which is the partial correlation coefficient squared) and measures the proportion of variations in the dependent variable (Y) explained by variations in another independent variable (say X_1), with the other independent variables held constant. However, the meaningfulness of this measure is dubious for practical work because, as often happens, the independent variables are themselves somewhat correlated and hence the several coefficients of partial determination will add to more than unity (i.e., 100 per cent) thus indicating overlapping effects.

---

[4] See R. J. Foote and J. R. Ives, "The Relationship of the Method of Graphic Correlation to Least-Squares," *Statistics and Agriculture* 1, U.S. Department of Agriculture, pp. 13–18; also, the bibliographical note below.

Conclusion

The following concluding points may be noted.

1. The coefficient of multiple determination indicates the proportion of variation in Y due to the *combined* effects of the independent variables. Graphically, an approximate indication of this is revealed by the closeness with which the estimated (calculated) values equal the actual values, as in Chart C.

2. An approximate indication of the relative importance of the independent variables in affecting the dependent variable is also given visually, as in Charts A and B, by the amount of scatter (i.e., degree of correlation) in the separate charts after completing the approximation process. Graphically, therefore, we can say that if the partial correlation between Y and X_1 is large and that between Y and X_2 is small, then, in general, X_1 has a greater effect on Y than does X_2. Similarly, if the correlations appear about the same, it could be assumed that both exert about an equal effect on Y. The regression line, therefore, shows the net effect of each of the independent variables on the dependent variable.

3. The slope of the regression line, i.e., the change in the vertical distance per unit of change in the horizontal distance, or $\Delta Y/\Delta X$ where Δ (delta) means "the change in," shows the *amount* of change in the dependent variable associated with a unit change in the corresponding independent factor after the influence of the other independent factors have been allowed for statistically.

Remark. If the slope is measured in percentages, it reveals the *partial elasticity*, which is defined as the percentage change in the dependent variable resulting from a 1 per cent change in the independent variable, with the other independent variables held constant. If Charts A and B had been plotted on double logarithmic paper, these elasticities could be determined directly from the chart. (Illustrations are given in Chapter 5 below.)

SIMPLE AND MULTIPLE RELATIONS

In the previous chapter, a brief explanation was given as to the distinction between simple and multiple correlation. Since the average student in economics or business rarely goes beyond an elementary course in statistics, and since such courses rarely go beyond simple correlation, a somewhat fuller explanation of the distinction will be useful in view of the fact that correlation analysis as a whole plays such an important part in economic measurement. As in the previous section the discussion will be essentially nonmathematical; emphasis will be placed on some of the important fundamental concepts rather than the computation technique as such.

Simple Relations

Where only two variables are involved, a dependent and an independent, the procedure for correlating them is that of *simple* correlation. The conceptual nature of the relationship can be expressed in words as "Y is a function of X," or in equivalent symbols in the form $Y = f(X)$.

If the relationship between the two variables is linear, the statistical process for arriving at the equation of relationship (also called the "estimating equation" or the "predicting equation") is known as *simple linear correlation.* The mathematical method for arriving at this equation—by first sketching a scatter diagram to see if the relationship appears linear, and then processing the data—is studied in most courses in elementary statistics. On the other hand, if the scatter diagram reveals a definite curvilinear relationship, the statistical process is known as *simple curvilinear correlation*, but the objective is still the same: to arrive at the equation which seems best to explain the nature of the relationship. There are thus a variety of possible curve types, and hence equations, that can be used to define the relationship between the variables. In practical business research, however, many of the common ones can be defined by a small number of rather simple equations. In each case the result is what is known as a *type equation*, and each type equation represents an entire *family* of curves.

Figure 3–2 presents six families of curves frequently encountered in economic measurement and forecasting, along with their corresponding

FIGURE 3–2

COMMON CURVE TYPES AND EQUATIONS

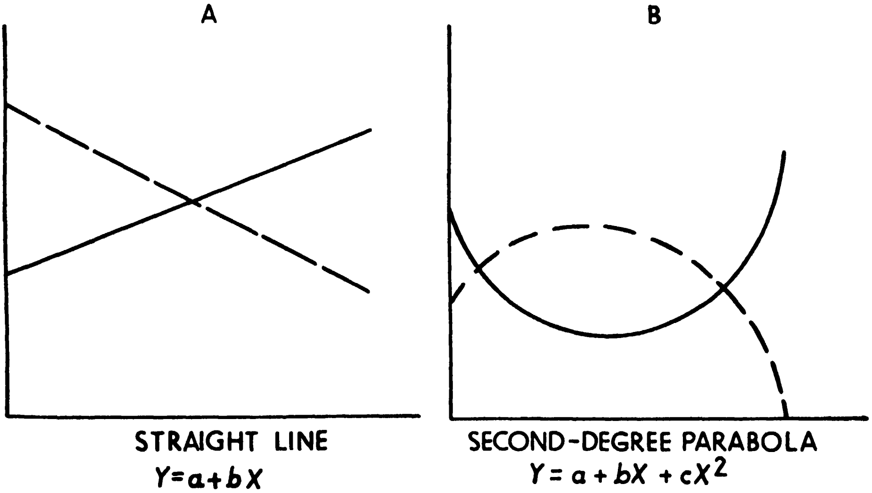
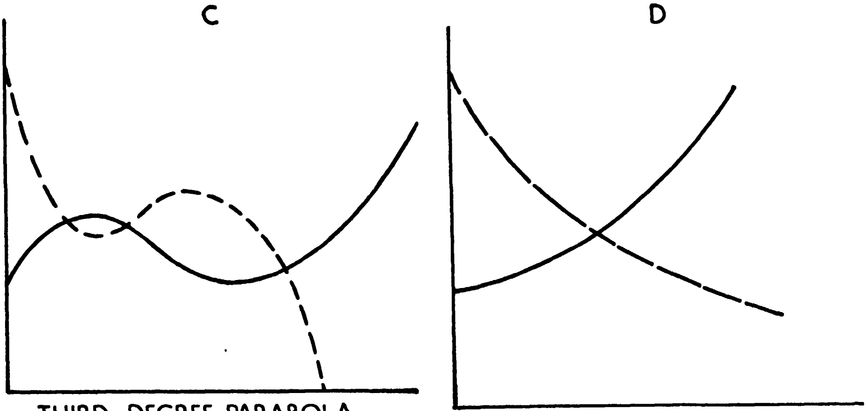


FIGURE 3-2—*Continued*



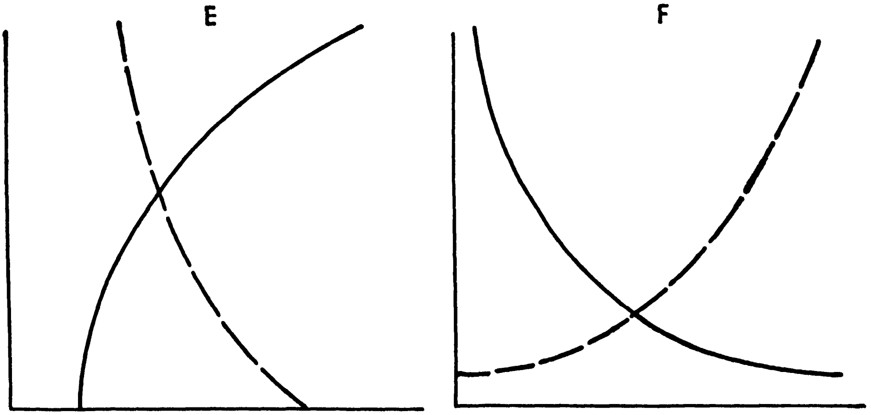
THIRD-DEGREE PARABOLA

$Y = a + bX + cX^2 + dX^3$

SEMI-LOGARITHMIC CURVE

$Y = ab^X$

OR $\log Y = \log a + X \log b$



LOGARITHMIC CURVE

$Y = aX^b$

OR $\log Y = \log a + b \log X$

RECIPROCAL CURVE

$1/Y = a + bX$

OR $Y = \frac{1}{a + bX}$

type equation. The capital letters in the equations represent the variables while the lower case letters are parameters. The shape of the curves, i.e., whether they follow the solid or dashed pattern, will depend upon the sign (plus or minus) of the parameters. In any event the basic forecasting problem is to find by statistical procedures the values of the particular parameters for the particular curve of the family which seems to fit the data best. For most forecasting problems the curves are ordinarily fitted by the method of least squares.[5]

---

[5] The reason for this is that if the residuals are normally distributed, this method yields a *probable* best fit as compared to any other curve in the same fam-

Multiple Relations

In many problems it is common to encounter situations in which the variations in the dependent variable can be more fully explained by the inclusion of more independent variables in the equation. When two or more independent variables are employed, the process is known as *multiple correlation.* Conceptually, the functional relationship is expressed symbolically as $Y = f(X_1, X_2)$ where two independent variables are concerned, or further X's may be included to represent additional controlling factors as in the equation $Y = f(X_1, X_2, X_3, X_4, \text{etc.})$. In the graphic method of multiple correlation described in the previous section, only two independent variables were employed. A closer fit of the calculated or estimating line might have been obtained, however, had more variables been included in the analysis.

The purposes of multiple correlation and the principles of measurement are much the same as those for simple correlation. Ultimately, a regression equation is fitted to the observed relationships on the basis of which predictions are made. An aggregate measure of correlation is also derived between the dependent and independent variables, the square of which gives the coefficient of multiple determination and denotes the proportion of total variations in Y explained by the X's in the equation. Measures of partial correlation and of partial determination can be derived as well, which indicate the degree of association and proportion of variation, respectively, between the dependent variable and any number and combination of independent variables in the regression equation that is desired.

As in simple correlation, multiple relationships may also be linear or curvilinear. A linear relationship between the variables in the regression equation is called *linear multiple correlation,* and exists when the scatter diagrams between the dependent and each of the independent variables, i.e., the Y-X_1 relationship, the Y-X_2 relationship, etc., appears to be best represented by a straight line. This was the type of relationship assumed in the earlier discussion of the graphic method. The regression or predicting equation would, for two independent variables involving linear multiple correlation, take the form

$$Y = a + bX_1 + cX_2,$$

and for three independent variables

$$Y = a + bX_1 + cX_2 + dX_3.$$

Each of the X's represent different independent variables while Y, of course, is dependent. If a fourth independent variable were included, the equation would add on the term eX_4, and for a fifth, fX_5, and so on. In-

---

ily. Basically, the least-squares method chooses the value which minimizes the sum of the squares of the deviations from the chosen value. Another common method, that of maximum likelihood, chooses as the estimate the particular value which maximizes the probability density. Both methods are illustrated in later chapters.

cluding further variables improves the ability of the equation to predict changes in Y due to variations in each of the X's. The value of a is the constant term in the equation and is equal to zero when the estimating line passes through the origin. The remaining coefficients b, c, d, etc., represent the rate of change in the dependent variable per unit change in each of the independent variables with the other independent variables remaining constant. Hence they are commonly called the *coefficients of net regression*, to distinguish them from *coefficients of gross regression* in simple correlation where no allowance is made for indirect influences on the regression. In econometric studies, as will be shown in later chapters, the object as in simple correlation is to arrive at the most probable value of these parameters, thereby enabling probability forecasts to be made based on past relationships.

Although linear multiple correlation is frequently encountered in many if not most practical problems, curvilinear relationships are common enough to warrant at least a brief statement. In the above linear regression equation, the value of Y changes at a constant rate with respect to changes in each independent value. In graphic terms, the regressions Y-X_1, Y-X_2, etc., are straight lines on the separate scatter diagrams; in mathematical terms, the regression equation involves only the first powers of the independent variables. It may happen, however, that Y changes at increasing and/or decreasing rates with respect to each of the X's. Graphically, this means that the regression lines on the separate scatter diagrams are curved rather than straight; mathematically, the regression equation would involve powers greater or less than 1 for the independent variables. Analyses of this type where the dependent variable exhibits a curvilinear relationship with one or more of the independent variables are called *curvilinear multiple correlation.* The principles are the same as with the linear case, except that the calculations are more complex and laborious. Some examples of curvilinear multiple regressions with four independent variables are the following:

$$Y = a + bX_1 + cX_2^2 + dX_3 + eX_4^2$$

$$Y = a + bX_1 + cX_2^2 + dX_3^3 + eX_4^4$$

$$Y = a + bX_1 + c\sqrt{X_2} + dX_3 + eX_4^2$$

From an ideal standpoint, therefore, the problem both in simple and multiple correlation is to choose the function that best represents the relationship according to what would be expected from economic theory. In reality, however, the underlying theory is often inadequate and the best that can be done is to choose the function that fits the data most closely. The significance of this will become evident in later chapters where it will be seen that even though economic theory might dictate a particular type of curve, the actual data appear better represented by some other type of curve quite different from that deduced in theory.

CORRELATION ANALYSIS IN FORECASTING

Correlation analysis is one of the most powerful tools of measurement available to researchers in applied science, and occupies a primary position in the kit of analytical techniques available to professional researchers in business economics. Before employing the method, however, the analyst himself must make a number of technical decisions based on *a priori* knowledge or considerations. These considerations constitute the basic problem areas involved in utilizing correlation procedures for forecasting purposes in business administration. A few of these more important problems and some indications as to their resolution are sketched briefly in the paragraphs below.

Choice of Equation or Curves

Once the relevant variables for a demand study have been selected, the analyst must make a decision as to their relationship. Mathematically, this means that the type of equation or system of equations to be used must be determined. Various curve types commonly used in forecasting have already been discussed. Suffice it to say, therefore, that the choice of a correct curve, e.g., straight line, parabola, or exponential, should be based on logical considerations about the variables, particularly since extrapolations are to be made beyond the range of the original data.

A consideration which is helpful in deciding on the choice of curves is whether the effects of the separate independent variables on the dependent variable are additive or multiplicative. Frequently, a multiplicative relationship can be transformed without much effort into an additive one by using either logarithms, reciprocals, roots or powers, or logarithms of logarithms. For example, consider the following multiplicative relationship which is the well-known compound interest curve of Figure 3–2D:

$$Y = ab^X .$$

If the variables are converted into logarithms, the following additive relationship is obtained:

$$\log Y = \log a + X \log b .$$

In this form, the analysis could be run by the graphic method and straight lines used, after first transforming all of the variables into logarithms. Similar illustrations of how multiplicative relations can be made additive may be cited, but they are unnecessary. The important points to realize are that (1) the choice of curves should depend as far as possible on logical economic considerations about the data, and (2) complicated curvilinear relationships can often be transformed into linear form, thereby reducing greatly the amount of work involved as well as enabling management to comprehend the results more easily.

In recent years the Cowles Foundation for Research in Economics has proposed that many analyses employing only single equations should be solved by using systems of equations or simultaneous equations. Their argument is that when single equations are used, the assumption is made that the current values of the independent variables affect the current values of the dependent variable, so that the cause-and-effect relationship is strictly one way in nature. Usually, however, this assumption is too limited and unrealistic according to the Cowles Foundation. They argue that, in reality, the structural variables in the economy are so interrelated that the cause-and-effect relationship is actually two way in nature. Thus, variations in the independent variables cause variations in the dependent variable which in turn cause variations again in the independent variables, and these actions are especially true when the data all fall within the same time period. Accordingly, they suggest that a prediction model be composed of a system of equations rather than a single equation, with each equation expressing an essential set of relationships between key variables. Simultaneous solution of the equations would then yield results that are more valid for forecasting purposes.

A simple illustration of this taken from elementary economics may serve to convey the concept somewhat more clearly. In the theory of pure competition, the economist sets out to determine or predict the price and output of a commodity under particular demand and supply conditions. Geometrically, the prediction is arrived at by drawing a demand and a supply curve, each expressing a particular price-quantity functional relationship, on a graph, and the *simultaneous* solution of these curves which occurs at their point of intersection then denotes the predicted price and quantity. Mathematically, the procedure is conceptually the same, but the equations for the curves are solved simultaneously instead. Although this example is an extremely simple one, economics abounds with such situations, usually of a more complex nature. Accordingly, econometricians at the Cowles Foundation, in order to make their analyses conform to these situations, have preferred to employ systems of simultaneous equations instead of single equations in their prediction models. The systems of equations are called "structural equations," and each structural equation is assumed to describe an economic relationship exactly except for random shocks.

Although the simultaneous-equation method has gained increasing use compared to the single-equation method in econometric forecasts of the total economy, the single-equation method is used almost solely by those business firms that employ econometrics in their own forecasting. There are several reasons for this, at least two of which may be noted.

1. *Relevance.* For the most part, where business forecasts are concerned, the cause-and-effect relationship is usually one way in nature, so that the single-equation method is valid. For example, in forecasting de-

mand, which is where econometrics has its widest use among firms employing it, the independent variables chosen are for the most part external to the firm, and they affect the company's demand but are not in turn affected by it. Thus, disposable income, the Federal Reserve Index of Industrial Production, and other measures of total economic change will typically enter as independent variables in a firm's demand forecast, because they have a bearing on that demand, but the latter in turn will have no appreciable effect on them.

2. *Adjustment.* The intercorrelated variables can frequently be combined into single independent variables so that intercorrelation becomes insignificant. Thus, as explained in the previous chapter, the price of food and the cost of living, both highly intercorrelated, can be combined into a single index and this new synthetic index used as an independent variable. Single-equation models are quite adequate, therefore, for most business forecasts, as will be seen in later chapters.

First Differences vs. Actual Data

The customary method in forecasting is to employ the actual data themselves as a basis for arriving at the forecast equation. Once the equation is derived and the values for the independent variables established, the forecast itself is an automatic result by simply substituting the known independent variables in the equation and then solving for the dependent variable. For example, the analysis may result in the forecast equation:

$$\text{Sales} = a + b\ (\text{disposable income}) + c\ (\text{advertising outlay}),$$

or in symbols, if S denotes sales, I is disposable income, and A is advertising expenditures, then

$$S = a + bI + cA\,.$$

In this equation, S is the dependent variable that is to be forecast; a, b, and c are the constants (parameters) of the equations, and the purpose of the mathematical method of correlation analysis is to arrive at their probable values; and I and A are the independent variables which, once they are predicted, are substituted in the equation, combined with their corresponding known constants, and the result is a forecast of S. Thus, suppose for illustrative purposes, the analysis were performed by the mathematical method and the constants of the equation were found to be $a = 10{,}000$, $b = .0000003$, and $c = .70$. The equation would then be written

$$S = 10{,}000 + .0000003I + .70A\,.$$

To forecast S by this equation, it is first necessary to predict the independent variables I and A for the same period. For instance, to forecast S for 1963, suppose, to keep the arithmetic simple, it is predicted that disposable income in that year will be $400 billion and that the company will spend

$100 thousand on advertising. The forecast is then a simple mechanical operation:

$$\begin{aligned} S_{1963} &= 10{,}000 + .0000003\ (400\ \text{billion}_{1963}) + .70(100\ \text{thousand}_{1963}) \\ &= 10{,}000 + 120{,}000 + 70{,}000 \\ &= \$200{,}000 \end{aligned}$$

Instead of specifying a particular year, say 1963, the forecast equation can be generalized for any time period; thus,

$$S = a + bI_t + cA_t \qquad (1)$$

where the subscript, t, refers to the time in question and can represent any year desired.

The method just described is fairly typical of the forecasting procedure involved when the actual data are used. It is sometimes useful, however, to employ not the actual data, but the *absolute changes* in the data from the preceding year (first differences) or perhaps the *percentage changes* in the data from the preceding year (link relatives). When the first differences are used, for example, the mathematically derived function can be employed to forecast the *change* in sales rather than the level of sales. By using a lagged subscript, S_{t-1}, to represent sales in the last time period as compared to sales in the current period, S_t, the forecast change in sales (ΔS) could be expressed

$$\Delta S = S_t - S_{t-1}\,.$$

Therefore, since

$$S_t = a + bI_t + cA_t \quad \text{(from equation 1 above)}$$

and

$$S_{t-1} = a + bI_{t-1} + cA_{t-1}$$

then

$$\Delta S = S_t - S_{t-1} = (bI_t - bI_{t-1}) + (cA_t - cA_{t-1})$$

or

$$\Delta S = b(I_t - I_{t-1}) + c(A_t - A_{t-1})\,.$$

To forecast the change in sales, it is only necessary to insert in the last equation the required values of I and A and then solve for ΔS. The forecast of the sales level in time t would then be the sum of actual sales in the preceding time period and the forecast change in sales from time $t - 1$ to time t. This result is not the same as would have been obtained by forecasting the level of sales directly from the original equation.[6]

What are the bases for distinction in using the actual data for the forecast as against using the first differences? It will be remembered that the first step in the graphic correlation method was to compute the average for the dependent variable. By using actual data, the analysis was ultimately able to show the relative importance of the independent variables in causing the dependent variable to change *from its average value* for the period of the analysis. Had first differences been used instead of actual data, the analysis would have shown the relative importance of the independent variables in causing the dependent variable to change *from one year to the next.* An advantage of using first differences became especially noticeable after World War II when postwar predictions were made on the basis of interwar data.[7] The structural changes that took place during and after the war brought sharp changes in the variables from their interwar average, making postwar forecasts grossly inaccurate. Studies based on first differences, however, were generally more accurate and often required little or no long-range extrapolations since they were designed to predict one year from the previous one. In addition to these considerations, the use of first differences instead of actual data may be preferable in the following instances: (1) If strong growth or trend factors tend to outweigh the more immediate effects of the variables, using first differences will tend to reflect the more direct period-to-period variations; (2) if intercorrelation is higher between the independent variables when actual data are used as compared to when first differences are used, the latter procedure is preferable; (3) if the residuals from the final analysis are serially correlated when based on actual data but not when based on first differences, the latter method is again preferable. By "serial correlation" is meant that the successive observations of a single variable, such as sales, are not completely independent of one another. Sales in one year are at least partially affected by sales in a previous year and will partially affect sales in the following year. Hence, a close relationship between the dependent and an independent variable may be due as much to the serial correlation within each series as to the causal relationship itself.[8] In other words, where time series are involved, the different observations are not randomly distributed and hence there is no logical basis for estimating the reliability of the coefficient of correlation.

---

[6] *Cf.* R. Ferber, "Sales Forecasting By Correlation Techniques," *Journal of Marketing* (January, 1954), p. 221.

---

[7] Until World War II, most analyses were based on actual data, although Henry Moore had advocated the use of first differences and link relatives as early as 1914. R. Ferber, *A Study of Aggregate Consumption Functions,* N.B.E.R., Tech. Paper 8, 1953, is worth examining.

[8] A possibility suggested by Ferber is to incorporate the serial correlation directly into the equation by using last period's sales as an extra independent variable. Thus

$$S_t = a + bI_t + cA_t + dS_{t-1}$$

The use of last period's sales in this manner reflects the effects of past influences as a whole on future sales levels. See Ferber, "Sales Forecasting," *op. cit.,* p. 227; also, the reference to Thomsen and Foote in the bibliographical note below.

The above considerations should not be construed as meaning that actual-data analyses are always inferior to first-differences methods. The latter procedures, for instance, would be difficult to use for longer-range predictions where extrapolations must be made much beyond the analysis period for which data are available, because they are most suitable for year-to-year predictions. Both approaches have their usefulness, and the choice of either one depends on the significance of the sort of issues raised above.

Time Lags and Time Trend

As pointed out in the previous chapter, one of the chief problems of forecasting is to secure a relationship that will remain in conformity with cyclical influences. The forecast formula may backcast excellently for the analysis period, but break down completely in predicting the future course (and particularly the turning points) of the dependent variable. One common reason for this is that the variables used in the formula all relate to the same time unit, e.g., current sales are a function of current income and current advertising expenditures. The introduction of a time lag may thus improve considerably the accuracy of the predictions. Time lags are used whenever the *effect* of a given independent variable manifests itself in a later time period, as, for example, when the assumption is made that advertising this year affects sales next year because it takes the average consumer that long to adjust his buying and for the message to "soak in" and take effect. Ideally, as many time lags as possible should be used for the independent variables, thereby reducing the uncertainty inherent in predicting the latter.[9] The length of the time lag chosen, however, should be based on logical considerations, because, statistically, the correct time lag to choose is not necessarily the one that makes the correlation coefficient a maximum.[10]

A second aspect of the time element in correlation problems is its customary usage by analysts as a catchall for the many factors known to change over time, e.g., tastes, technology, etc., but for which data are unavailable or which, singly, would be expected to have a small effect on the dependent variable. The result of this is that "time" is often incorporated in the function as an independent variable[11] with the recognition that it is a substitute for other causal variables that are being omitted. Some analysts, however, use an alternative approach of excluding time from the original analysis and then checking the final residuals to see if they exhibit a time trend. If they do, either another variable can be added which explains this trend, or, if the necessary data are unavailable, "time" as such can be added as an additional variable. Hence, when a time trend is used, an attempt should be made to explain it on logical grounds, although sometimes this may also be difficult to do. Otherwise, recognition should be given to the dangers inherent in this source for making predictions.

---

[9] An approach suggested by Duesenberry and Modigliani is to lag income by its past cyclical peak, on the assumption that, in forecasting consumption, it is more difficult for people to adjust their standard of living downward than it is to adjust it upward. Therefore, the future spending and savings habits of people are at least in part conditioned by their highest past level of income and their living standards at that time. (J. Duesenberry, *Income, Saving, and the Theory of Consumer Behavior;* and F. Modigliani, "Fluctuations in the Saving-Income Ratio: Problem in Economic Forecasting," in N.B.E.R., *Studies in Income and Wealth,* Vol. 11, pp. 371–443.)

[10] This may seem surprising, but is nevertheless true. (See J. Marschak, "Economic Interdependence and Statistical Analysis," *Studies in Mathematical Economics and Econometrics,* pp. 135–50, especially p. 150, note 19.)

Per Capita and Deflated Data

In demand analyses, particularly where elasticity coefficients are being estimated, there is no consistent agreement as to whether these estimations should be based on total or per capita data or on deflated or undeflated price and income series. There is, however, the obvious consideration that the several series should be consistent with one another. Thus, if per capita consumption data are being forecast, the demand shifter, such as national income or industrial production, should also be expressed in per capita terms. The use of per capita data, however, requires some thought as to the population groups that are to be included as a divisor. In forecasting the demand for cigarettes, for example, only adults would be included in the analysis, and perhaps shifting weights could be ascribed to both men and women. Adjustment problems of this nature are discussed later in Chapter 5 with respect to demand measurement.

Intercorrelation

The problem of intercorrelation among the independent variables has already been mentioned at several earlier points in the discussion of graphic correlation. It should be stated now that high intercorrelation, i.e., a high degree of correlation between one or more independent variables, has a more serious effect on the graphic method as compared to the mathematical method.

Specialists on the subject point out that a major contribution of the graphic procedure is a simplified method whereby the first approximations of the partial regression lines or curves are progressively improved. Further, each successive improvement (approximation) varies inversely with the degree of intercorrelation.[12] But where the differences between two successive approximations may be unnoticeable, the analysis may be terminated at too early a stage as compared to the regressions that would have been obtained by the equivalent mathematical procedure. In view of this, what might the analyst employing the graphic method do to offset this possibility? One alternative, as mentioned earlier, is to use first differences when the variables are highly intercorrelated in terms of the actual data. A second possibility, say when there are two independent variables, is to combine them as a ratio instead of using the two separate variables, if this is feasible. A third alternative is to try a substitute variable which may reflect the displaced variable indirectly, e.g., using railroad passenger travel instead of railroad fares as an independent variable in forecasting the demand for airline transportation. These are just a few of the dodges that can be employed, and the analyst can undoubtedly think of others, depending on the particular circumstances.

---

[11] Assuming its regression coefficient differs significantly from zero; if not, it may be omitted from the analysis.

[12] Foote and Ives, *op. cit.*, pp. 13–18, show this to be true. (See also the bibliographical note below.)

Conclusion: Forecasting Accuracy

In the previous chapter, some concluding comments were presented in the form of qualitative criteria for judging the effectiveness of a forecasting method. Regardless of the method used, however, every forecast that ultimately emerges must eventually be evaluated. What criteria exist for performing this evaluation? Evidently, the criteria themselves must be quantitative in nature, and at least three simple ones that are frequently employed may be mentioned.

1. The percentage deviation of the forecast value from the actual or realized value can be computed. Where several forecasts have been made for a succession of periods, the separate percentage deviations can be combined and averaged to obtain an average percentage deviation.

2. A set of forecasts can be evaluated by comparing them with a benchmark such as a naive model set of predictions. An example of the latter would be a prediction of next year's sales as a mere extension of this year's sales, or perhaps some multiple or fraction of this year's sales level. This provides an indication of the value of the present (presumably more elaborate) method compared to a more simple procedure.

3. To remove the spurious results that may be obtained by comparing levels of forecasts with actual sales, the criterion can be changed from one of comparison to one of direction of change. That is, because of the influence of serial correlation, a comparison of levels of forecasts with actual sales, as in the first standard above, will appear favorable, because sales each year have some influence on sales in the following year. By comparing directions of change rather than levels, the association between predicted and actual figures due to level is largely removed.

BIBLIOGRAPHICAL NOTE

Two good general sources as a basis for further treatment of the topics covered in this chapter are E. F. Beach, *Economic Models*, and R. Ferber, *Statistical Techniques in Market Research*. The first, especially in the beginning chapters, contains a clear exposition of the elements and nature of model construction with no particular knowledge of mathematics required; the second treats in greater detail the subjects of sampling, experiments, and correlation, with emphasis on practical applications rather than mathematical derivations, and can be readily handled by those with a year of statistics as a background. Also at the same level is an article by E. G. Bennion, "The Cowles Commission's 'Simultaneous Equation Approach': A Simplified Explanation," *Review of Economics and Statistics* (February, 1952), which provides a good account of the logic behind the use of multiple- rather than single-equation systems. A practical application of controlled experiments to merchandising problems is discussed in M. Brunk and W. Federer, "Experimental Designs and Probability Sampling in Marketing Research," *Journal of the American Statistical Association* (September, 1953). The method of graphic correlation is described by its originator, M. Ezekiel, in his *Methods of Correlation Analysis*, Chapters 6, 14 and 16, applied to both linear and curvilinear relationships. A clearly written and critical evaluation of the technique is developed in an article by W. Malenbaum and J. D. Black, "The Use of the Short-Cut Graphic Method of Multiple Correlation," *Quarterly Journal of Economics* (1937–38), pp. 66–112. Although the method is not explained in any of the recent editions of standard works in business or economic statistics, it has been used extensively by agricultural economists in forecasting commodity prices and is described along these lines in the following works: F. Thomsen and Foote, *Agricultural Prices;* W. Waite and Trelogan, *Agricultural Market Prices;* and G. Shepherd, *Agricultural Price Analysis.* In addition, Ferber's work mentioned above also illustrates the technique.

For those with little or no statistical training, the following are suggested: on sampling, J. Lorie and Roberts, *Basic Methods of Marketing Research*, contains several chapters on applied sampling which are most readable and informative; on controlled experiments, a brief exposition is given in M. Blair, *Elementary Statistics*, rev. ed., pp. 525–29, 601–3, with emphasis on the Latin square, and in a book of the same title by R. Sprowls, pp. 365–67; on graphic correlation, the texts on agricultural prices mentioned above are most adequate; and on the use of correlation analysis in forecasting, R. Ferber, "Sales Forecasting by Correlation Techniques," *Journal of Marketing* (January, 1954), and F. L. Thomsen and Foote, *Agricultural Prices*, chap. 16, are recommended. The last two cover essentially the same ground, while the latter contains an excellent exposition of the problem in general. Finally, a good comprehensive treatment of econometric model construction as it applies to forecasting the total economy is E. C. Bratt, *Business Cycles and Forecasting*, 4th ed., chap. x, which can be read by the nonmathematical student in place of more technical works such as Leontief's "Econometrics," in H. S. Ellis (ed.), *Survey of Contemporary Economics*, Vol. I, chap. 11.

QUESTIONS

1. (*a*) What has been the purpose of this chapter? (*b*) Why?
2. What basic methods are commonly employed in the measurement of economic phenomena?
3. (*a*) How does sequential sampling differ from the conventional fixed-size sample? (*b*) What advantages does sequential sampling offer over con-

ventional sampling methods? (*c*) In general, what characteristics should the data possess in order for sequential sampling to be preferable? Explain.

4. How does a controlled experiment differ from a sample survey?
5. What is a Latin square?
6. What dangers, if any, exist in using controlled experiments in demand studies?
7. Engineering methods of measurement have their greatest application in what areas of business management?
8. What is the purpose of correlation analysis?
9. In the earlier years of econometric science, the studies made dealt largely with agricultural goods rather than manufactured goods. Why?
10. What is the final objective in the mathematical method of correlation as compared to the graphic method?
11. (*a*) Why is it necessary when performing a graphic correlation that the regression line, if straight, pass through the means of its two variables? (*b*) Of what practical value is this rule? (*c*) What is the purpose of drift lines?
12. Usually, the dots on a scatter diagram do not fall exactly along the regression line, but instead are scattered about it. What does this indicate?
13. (*a*) Using the formula for the coefficient of multiple determination on p. 69, compute R^2 from Table 3–1, p. 59. (*b*) In one sentence, interpret your result. (*c*) How is your coefficient of multiple determination related to the dashed and solid line in Chart C of Figure 3–1, p. 60?
14. What are the economic implications of adding further variables to an econometric analysis?
15. (*a*) What relationship is indicated by the *slope* of the regression line, i.e., the change in the vertical distance per unit change in the horizontal distance, on a scatter diagram in an econometric analysis involving two or more independent variables? (*b*) What if the slope is measured in percentages?
16. (*a*) What rules should be kept in mind in choosing the type of equation or curve for an econometric analysis? (*b*) When might a simultaneous-equation model be preferred to a single-equation model?
17. (*a*) What are "first differences"? "Link relatives"? (*b*) In general, when is the use of first differences preferred over the use of actual data in forecasting?
18. (*a*) What is the reason for sometimes including a time trend in an econometric analysis? (*b*) Ideally, how should "time" be incorporated in the equation? Why?
19. Why are per capita and deflated data employed in some studies?
20. What techniques might be employed to eliminate the significance of intercorrelation and thereby facilitate the use of single- rather than simultaneous-equation models?
21. (*a*) What objective criteria may be employed in evaluating forecasting accuracy? (*b*) In the light of this and the previous chapter, what rigorous objective criterion for evaluating forecasting accuracy would you suggest that was not mentioned in the above? (*c*) What forecasting method that you know of would come closest to meeting your proposed criterion?

PART II

Adjustment to Uncertainty

Having outlined the uncertainty framework of management decision making and the general methods of forecasting available for reducing that uncertainty, we turn now to the problems of adjustment in the more specific areas of business management. Thus, the uncertainty areas considered will involve profit, demand, production, cost, pricing, competition, and capital management. Stress will be placed on the economic aspects of these problem areas particularly from a measurement, forecasting, and policy standpoint, so as to provide an intelligent basis for managerial action.

Chapter 4

PROFIT MANAGEMENT

Profit is the ultimate test of a firm's well-being and a comprehensive indicator of management's ability to fulfill its coordinative function of decision making and planning. Since the search for profit is, after all, the reason for a business firm's existence, this chapter, and all of the chapters which follow, is an analysis and treatment of the forces which determine profit or the lack thereof.

In this chapter we undertake a detailed examination of various profit aspects with a major view to relating the theoretical with the "practical." This means a discussion of various profit theories that have been put forth over the years, an analysis of profit measurement, a brief presentation of some techniques for profit control and prediction, and a reconciliation of the profit-maximizing principle of economic theory with the (nonmaximizing) profit policies apparently adopted by business firms.

PROFIT THEORIES

The history of the development of economic thought reveals an abundance of profit theories which, in varying degree, are based upon some one or a combination of the aspects of profit—whence it derives, the economic function it performs, to what productive factor or factors it is distributed. We will, therefore, present a brief summary of profit theories, not individually or comparatively, but in terms of the three major categories in which all profit theories may be more or less classified: (1) compensatory or functional theories, (2) friction and monopoly theries, and (3) technology and innovation theories. This classification is not all inclusive, nor does it imply that particular theories may not contain elements of others. It merely points out the different lines that have been followed historically in the course of thinking on the subject and represents a logical arrangement of ideas for approaching the problems of managerial decision making.

Compensatory or Functional Theories

This group of theories holds that economic profits (surplus) are the necessary payments to the entrepreneur in return for the services he performs in coordinating and controlling the other productive factors. It is the entrepreneur who organizes the factors of production into a logical sequence, combines them efficiently, establishes policies and sees that they are carried out, and in other ways acts in both a coordinating and supervisory capacity. Profits, therefore, are his compensation for fulfilling these functions successfully. In like manner, losses are the penalty for unsuccessful entrepreneurship.

This group of theories, propounded in the early nineteenth century and represented later in the United States mainly by the economist Francis Walker, placed the entrepreneur in the position of a higher type of laborer, or at least made him synonomous with the individual proprietorship type of business enterprise. When attempts were made to apply the theory to the modern large corporation with its separation of ownership and control, the results appeared confusing and contradictory. In this form of business organization, the coordinative function is usually delegated by the owners (stockholders) to salaried executives. If the latters' remuneration is taken to be "profits," despite its contractual form, the theory still leaves unexplained the residual income of the enterprise that goes to stockholders who exercise no active control. The only alternative, if these theories were to be consistent with their original definition, was to allocate a share of the entrepreneurial function to stockholders. But attempts to do this are not in accord with reality where the corporation is an organization of active leadership by managers and passive ownership by stockholders. With the growing importance of the large corporation as a dominant type of business organization in the American economy, this set of profit theories lost their significance and a group of "friction theories" emerged in their place around the turn of the century.

Friction and Monopoly Theories

By the year 1900, the theory of a stationary economy was well on its way toward becoming a complete and unified system of thought. Against this setting the noted American economist J. B. Clark constructed an economic model that was intended to be a reconciliation between static laws of theory and the dynamic world of fact. According to stationary theory (or perfect competition as it is commonly called today), the economy is characterized by a smooth and frictionless flow of resources, with the system automatically clicking into equilibrium through the free play of market forces. Changes may occur that will occasion a departure from equilibrium, but so long as resources are mobile and opportunities are equally accessible to all economic organisms (i.e., knowledge is perfect), the adjustment to change and a new equilibrium will be accomplished quickly and smoothly. In this type of economic equilibrium all factors of production would receive their opportunity costs; the revenues of every enterprise would exactly equal its costs (including the imputed wages and interest of the owner), and hence no economic surplus or profit residual could result. In the real world, however, such surpluses do occur, and in accordance with the theory they can only be attributed to the frictions (or obstacles to resource mobility) and changes that actually characterize a dynamic economy. In the long run, according to the theory, the forces of competition would eliminate any surpluses, but the surpluses continue to recur because new changes and new frictions continually arise. Profits, therefore, in contrast to the earlier compensatory theories outlined above, are not attributable to any particular function; they are the result of institutional rigidities in the social fabric that prevent the working out of competitive forces and are to the temporary advantage of the surplus recipient. Such imperfections in resource fluidity made it possible to generalize about "unearned" incomes: rising rents attributed to enhanced values of limited land resources and the natural pressure of growing population and increasing urbanization (Henry George); "abnormal" profits ascribed to the existence of monopolistic and even exploitative elements of a favored capitalist minority (a modernized restatement of Marx); and, in fact, all surpluses preempted by owners of any resources (including labor, capital, and managerial skills) because of the institutional frictions of an otherwise fluid system (Veblen and Hobson).

There are many illustrations from real life to substantiate the friction theory as a cause of economic surplus. The construction of military posts during the war brought profit bonanzas to neighboring cities; the Suez crisis (fall of 1956) rescued the domestic oil industry from what was threatening to become a highly embarrassing oversupply of refined oil products; the existence of patents and franchises enables many firms to reap profits by legally excluding competitors from the field; a favorable location for a business may result in the value of the site exceeding the rental payment for it; or, in general, the control of any resource whose supply is scarce relative to its demand, provides a basis for pure or windfall profits. In such instances a surplus would not arise if resources were sufficiently mobile to enter the market, or if the economy were frictionless (perfect) in its competitive structure. At best, if any surpluses did arise, they would be short-lived and would vanish entirely when the adjustments had time to exert their full effect in the market. But social processes—customs, laws, traditions, etc.—make these rapid adjustments impossible.

Technology and Innovation Theories

An innovation theory of profits can be developed which, when cast into an uncertainty framework, probably goes further than any other theory toward explaining in a realistic way the historical development of business enterprise.[1] An innovation is defined as the setting up of a new production function. A "production function" is the physical relation between the output and various kinds of inputs (capital, land, labor, etc.) in a production process.

---

[1] The innovation theory as originally expounded by the late Professor Joseph Schumpeter was an attempt to explain business cycles, not so much the causes and distribution of profits.

From a broad business standpoint, an innovation may embrace such a wide variety of activities as the discovery of new markets, differentiation of products thereby yielding wider consumer acceptance, the development of a new product, or in short, a new way of doing old things or a different combination of existing methods to accomplish new things. There is an important distinction to be made between invention and innovation. Invention is the creation of something new; innovation is the adaptation of an invention to business use. Many inventions never become innovations.

The original purpose of the innovation theory as propounded by Schumpeter was to show how business cycles result from these "disturbances" and successive adaptations to them by the business system. He never stated (as is sometimes implied) that innovation alone was the cause of change or disturbance in the economic system. His procedure was to assume a stationary system in equilibrium—in which all economic life is repetitive and goes on smoothly, without disturbance. Into this system a shock—an innovation—is introduced by an enterprising and forward-looking entrepreneur who foresees the possibility of extra profit. The quietude and intricate balance of the system is thus shattered as if invaded by a Hollywood-staged cattle stampede. The successful innovation causes a herd of businessmen (followers rather than leaders) to plunge into the new field by adopting the innovation, and these mass rushes create and stir up secondary waves of business activity. When the disturbance has finally ironed itself out, the system is settled in equilibrium once again, only to be disturbed later on by another innovation. Economic development thus takes place as a series of fits and starts (cycles) rather than progressing as a smooth and continuous growth.

The manager who is considering the introduction of an innovation must subjectively forecast the effect of that innovation on expected profits. The expected profit is the sum of expected receipts less the sum of expected expenses at all moments of time within the economic or planning horizon. This horizon is the length of time over which managers plan economic activity. If future sales and costs (and hence profits) were known with certainty, the span of the planning horizon would be infinite. But in a world of uncertainty where forecasts must be subjective, the time length of the planning horizon will differ among managers and will depend on the extent to which they formulate effective expectations and plans in a temporal vein. "Effective" expectations, therefore, are expectations that are held by managers with a degree of "subjective certainty" sufficient to cause action or the establishment of a plan (as would occur, for example, if discounted returns exceeded costs).[2]

---

[2] Thus, in an uncertainty framework, an innovation may be defined as "such changes in production functions, i.e., in the schedules indicating the relation between

UNRESOLVED CONSIDERATIONS

The current state of profit theory leaves some questions still unanswered. Space does not permit a detailed examination of the problems and their many ramifications, but some reflections on at least two major issues are of interest: (1) the distribution of profits, and (2) the import of the innovation theory.

Distribution of Profits

Perhaps the most significant development in the history of profit theory has been the separation of ownership and control in the publicly held corporation. Prior to the emergence of the corporate form of business organization the entrepreneur performed the dual functions of owner-manager, thereby receiving an equitable claim on all residual income of the enterprise as well as exercising coordinative control over the firm's resources (decision making). But with the growth of the corporation, the stockholder assumed the ownership function while that of coordination was transferred to the domain of professional managers. This being the case, a problem arises as to what might be the "proper" allocation of a company's profits, after paying some "fair" or "reasonable" dividend to the stockholders. The following alternatives separately or in combination are possible: (1) the surplus can be paid to management, (2) it can be given to the general public in the form of lower prices, (3) it can be fully distributed to the stockholders, or (4) it can be plowed back into the business.

1. *To Management.* It may be argued that in a sense all of these alternatives are employed in some combination, though in varying degree. The use of bonuses and stock option plans for management is a means of compensating the latter for a job well done, and together with the regular salary paid them probably constitutes something approaching an opportunity cost wage. Good management deserves handsome compensation, but since it may probably be reasonably assumed that management tends to take care of itself, and that competition for managerial talent does exist, it would seem unreasonable to conclude that excessive distributions of the "surplus" profit be made to management.

2. *To the Consumer.* The second alternative, unless carefully interpreted, is, on the face of it, ridiculous and unworkable. It requires business not merely to self-impose a restraint on its desire for profit (an attitude which will be discussed later in the section on Profit Policies), but to make

---

the input of factors of production and the output of products, which make it possible for the firm to increase the discounted value of the maximum effective profit obtainable under given market conditions." By market conditions is meant prices, or demand and supply schedules. Discounted expected prices and schedules as well as current ones are included. An increase in discounted maximum effective profit means an increase in the sum of surpluses of effective receipts over effective expenses. (See O. Lange, "A Note on Innovations," *Review of Economic Statistics,* Vol. XXV.)

the restraint operative at what might be a level which either eliminates the small firm in industries characterized by oligopoly, or discourages an adequate flow of equity capital, or prevents the grim reaper of competition from eliminating the inefficient producer. If this function is transferred to government, it means a radical change in the institutional structure of what we know to be a capitalistic economy. In a sense, the function is being performed by both business and government, but not to the degree suggested by the above-mentioned alternative. Thus, the enforcement of our antitrust policies discourages the leading oligopoly firms from lowering prices (benefiting the consumer) as a means of reducing "surplus" profits, for this would bankrupt the smaller competitors and increase the monopolistic position of the leading firms and, in turn, put them in the position to make (for themselves) still larger profits. Interestingly enough, the postwar experience in the automobile industry has been a gradual attrition on the part of smaller firms—as well as a desperate struggle by the number 3 company in the Big Three to keep its share of the market—and this attrition has come about in the face of rising car prices. It has reflected, in other words, a shift in consumer preference in favor of the two leading companies.

3. *To the Stockholder.* The last two alternatives are best discussed in conjunction rather than separately, for the decision to pay out a dollar in dividends implies a simultaneous decision not to plow back that dollar into the company. In either case, whether earnings are paid out or plowed back, the stockholder benefits, though, depending on his tax status and cash needs, one stockholder will prefer maximum payout while another will prefer maximum plow-back. The former gets his benefits in the form of immediate cash income; the latter in the form of capital gains and enhanced future earning power.

Evaluation of the various alternatives suggested seems to us to lead to the only clear-cut solution that is at all reasonable—the business has been organized for the benefit of the ownership element, and so long as it operates within the constraints imposed by society, any profits derived from such operation must be allocated in a manner which will best suit the owners. We have already indicated that within the ownership group there is likely to exist a conflict of interest in that high-income stockholders will want low dividend payouts, while lower-income and tax-free institutional stockholders will prefer high dividend payouts. If we concede that a corporation should properly be managed in the best interests of the stockholders, the question arises, in view of this possible conflict of stockholder interest in the publicly held corporation—what are the best interests of the stockholders?

The problem cannot be properly treated at this point, for dividend and earnings-retention policies must be established in conjunction with capital expenditure planning. The decision to plow back earnings, in other words, must give consideration to the company's capital investment program, and cannot be determined in a vacuum, as it were. We must, therefore, defer our conclusions on this matter to the last two chapters of this book in which a detailed examination is made of capital expenditure planning.

Status of Innovation Theory

As it stands, the innovation theory is a "great-man" theory of history and thus provides a useful hook on which to hang the development of business in capitalist countries. The innovating decision maker is here cast in the role of determining the intensity and pace of economic growth.

When conceived in its broadest business sense as a new way of doing things, the innovation theory can go a long way in helping to explain such great historical episodes as the rise of mercantilism and the industrial revolution itself, as well as the underlying structural changes that took place in American business during the latter half of the nineteenth century and which have been in continued evidence since then. In terms of present-day experience, the broader aspects of the innovation theory can be seen in everyday business life: the development of new markets (automatic drive in automobiles, and television), revolutionary products providing "better living through chemistry" (nylon and orlon), amazing therapeuticals ("mental drugs" and antibiotics),[3] new promotional methods (give-aways and quiz programs), new metals (titanium and zirconium), a revolutionary source of energy (atomic fission and fusion), and new fuels (lithium and boron). The innovation theory places stress on the dynamic (uncertain), ever-changing nature of capitalism. It points out quite vividly that the only limits to human progress are the inherent limits to human beings themselves. And now even this may not be a seriously restraining factor with the impending advent of the "automatic" factory of the electronic age.

From the standpoint of managerial economics, the value of a theory is not so much determined by how well it explains the past or even the present, but how well it can predict the future. For this purpose, the innovation theory is of little use. For since the scientific principles, technical know-how, materials, and skills—all the ingredients necessary to bring to the business world the reality of a new product, service, or method of production or of distribution—are, at any time, known and available long before the innovation bursts forth on the business world, why does the explosion take place when it does, and neither sooner nor later? On the surface, the answer is that an innovation takes place when the possibility for profit is recognized by someone willing and able to exploit the potentials he believes to be inherent in the opportunity he sees. Basically, however, the answer lies in the structural environment of complex underlying pres-

[3] How amazing and profitable a successful innovation can be is illustrated by the sales of the so-called "mental drugs" which from zero in 1954 went to about $85 million in 1955 and $147 million in 1956.

sures, institutional and otherwise, which bring the innovator on to the scene. For the innovation theory to be complete, it must be reshaped in terms which facilitate the prediction of the innovation and its ramifications. In this form the innovation theory would serve to reduce the uncertainty that is inherent in forward planning. And in its broader application such a theory would actually explain the entire course of economic development. This is really what Karl Marx attempted, and failed to accomplish, a century ago in building his theory of economic history. It appears, at least intuitively, that the innovation theory may provide the point of departure for a similar venture, in which case we would not merely have an innovation theory, but a *theory of innovations.*

PROFIT MEASUREMENT

The Allocation Problem

When it comes to measuring profit, the major difficulty is introduced by the requirement of allocating to a given accounting period the "correct" revenues and costs deemed to be attributable to that period as distinct from previous and subsequent periods. The "true" profitability of any investment or business operation (as will be made quite clear in our chapters on capital planning) cannot be determined until the ownership of the investment or business has been fully terminated, so that the need to measure profits over a particular segment of the total life span of the investment imposes a degree of arbitrariness which cannot be avoided. Although arbitrary allocation to a given accounting period is necessary with respect to both revenues and costs, it is the latter which has received the greatest amount of attention, particularly with respect to depreciation accounting and inventory valuation. Aside from the accounting aspects of cost (and profit) measurement, there are certain important economic considerations which we would like to deal with first.

Significance of Economic Cost

In the economic literature, much attention is paid to the possible discrepancies that might arise, in the determination of profit, out of the failure to account for all costs. The point is made that certain portions of accounting profit may actually include elements of cost and that it is important to recognize these economic costs, as well as the more obvious cash outlays, and such items as development costs and capital expenditures which are amortized over the future. There are, in short, three possible sources of discrepancy: (1) the entrepreneur's wages (which he could earn by working for someone else), (2) rental income on land employed in the business (which the owner could receive by leasing the property to another firm), and (3) a minimum or "normal" profit (which would be just enough to compensate the owner for his capital investment and which he presumably could earn by putting his money to work in somebody else's business at equivalent risk).

The above items are all deemed to be costs for the simple reason that an entrepreneur who failed to secure a net revenue at least equal to their total would, in the long run, withdraw from the business, hire himself out to another firm, lease or sell his property, invest his funds in some alternative undertaking, and improve his economic position. Thus arises the technical, economic meaning of cost—that minimum compensation necessary to keep a given resource or factor of production in its stated employment in the long run. Frictions and various other market imperfections will cause resources to remain in their existing employment at less than economic cost, and there are many situations in which resource owners receive compensation in excess of economic costs, but in the long run there is sufficient mobility of resources which tends to eliminate these discrepancies (under dynamic, real-world conditions, changes always occur to redistribute the discrepancies and introduce new ones).

Cost, in the economic sense, is thus viewed as a payment necessary to keep resources out of (the next most attractive) alternative employment, since a payment which is below economic cost will result in an eventual shift of the resource to the alternative opportunity—hence the term "opportunity cost."

Our specific concern in this section with the problem of profit measurement raises the question of how to deal with these potential discrepancies between accounting and economic profit, for it follows from the above that economic profit or "true surplus" is equal to accounting profit less certain unaccounted for costs. Actually, the discrepancies are not too serious as applied to the corporation, and are most likely to exist in the small proprietorship. Thus, in the corporation, management is hired and receives, presumably, an opportunity cost wage. These wages are treated as expenses, along with the wage payments to all employees, and are deducted in determining final profit. Properties are ordinarily rented, and these rents are deductible in determining the final profit. To the extent that real estate is owned rather than rented, it is frequently segregated into a special real estate or building corporation subsidiary from which the property is "rented." Where the latter device is not used, the real estate is treated as part of the total investment which the firm seeks to employ profitably (the rental value being readily determinable). We are, therefore, left with this one possible source of discrepancy between accounting and economic profit—the earnings on the invested capital.

A corporation derives its long-term capital from any one or a combination of three external sources: the sale of bonds, preferred stock, and common stock. The bondholder's contribution is obtained, however, at an opportunity cost interest rate, and this cost is recognized in determining profit. As for the preferred stockholder, while he is legally an owner so that profit is computed before the distribution of the preferred dividend, his position is really that of a "limited partner." Thus, from the point of view of the common stockholder, the preferred dividend constitutes an opportunity cost payment for the use of the preferred stockholder's capital. This is, of course, objectively determinable, and in fact is always deducted in determining "net profits available for the common stock." We are then left with only one significant element of discrepancy between accounting and economic profit: the cost for the use of the common stockholder's contribution (including reinvested earnings) to the corporation. This "normal profit" on the stockholders' capital is the amount by which accounting profit exceeds economic profit in the corporation. This element is, furthermore, measurable—it is the amount that would be earned elsewhere on investments of equivalent risk, and unless the existing investment process is capable of producing this opportunity rate of return, the capital will be gradually withdrawn from its employment in search for greener pastures.

Problems in Measuring Accounting Profit

In addition to errors arising from a failure to give consideration to certain economic costs, there are the even more serious errors which arise from the accounting techniques themselves. The difficulties are not due to the failure of the accounting profession to produce the right techniques. They arise simply out of the fact, as stated earlier, that the true profitability of an investment cannot be precisely determined until the process has been terminated, and that for any period other than the full life of the investment profits can only be estimated, which in turn means that revenues and costs must, to some extent, be arbitrarily allocated to the period in question.

However, various factors impel the periodic determination and reporting of profits. Stockholders wish to know how their investment is faring; the government wants its taxes; and management needs a guide for future decision making and a measure of the success (or lack of it) of past decisions made. Thus, despite the dilemma that exists, the bull must be grasped by the horns. In doing so we will attack the problem from the economist's point of view.

The accountant, at least for legal reasons if for no other, is primarily concerned with historical fact so that profit is to him an *ex post* concept based on past transactions. The economist views profit as a surplus in excess of all opportunity costs, so that past outlays are of only partial significance, for the cost allocations arising from these past transactions must be modified by current facts. To state this in more concrete terms, the economist would say that the profit earned in period "A" is equal to the growth in value of the enterprise from the beginning of the period to the end of the period (after adjusting for any distributions by, or contributions to, the firm during the period). This increase in value is a reflection not only of what we ordinarily understand to be the results of operations during the period, but of changes in asset values (plant, equipment, inventories) as well. Thus *profit, in an economic sense, would be the difference between the cash value of the enterprise at the beginning and end of the period.*

We have thus laid a base from which we can proceed to evaluate certain accounting conventions used in arriving at accounting profit. The two major areas in which discrepancies are most likely to be produced, and which will receive detailed consideration here, are: depreciation accounting, and the significance of price-level changes on asset valuation.

Depreciation

In carrying on business activity, the firm's buildings, machines, and other equipment wear out with time and use, so that eventually a company's entire investment in such assets becomes worthless. In order, therefore, that the corporation's income be properly stated and that the cost, less salvage value, be recovered by the time the assets are abandoned, the accountant makes as a charge against annual income the amount of the decrease in value during that period. This charge is called depreciation and is usually prorated in equal amounts over the life of the asset. For this reason it is called the "straight-line method."[4]

Stated thus, simply, there would seem to be little more to be said about the subject, except to recognize that the importance of this operating cost to the enterprise would vary widely from one company to another, depending on the composition of the business assets. Thus, characterized by extremely large depreciation charges are such companies as those engaged in steelmaking, railroad and airline transportation, chemical processing, and the production of primary aluminum; while insurance companies, banks, investment funds, and advertising and merchandising establishments bear relatively small depreciation costs. The subject, however, does not end with this simple observation, for it is complicated by controversy over the true function of depreciation, the proper method for measuring it both for purposes of reporting net income to stockholders and taxable income to the government, and (a recent development) its proper use as a tool for stimulating capital formation and directing investments along lines deemed to be in the national interest.

Measuring Depreciation. While the straight-line method has been by far the most widely used in industry, the depreciation pattern has undergone substantial change in recent years because of the federal government's desire to stimulate and direct investment in new plants and equipment. This was first done by permitting, for tax purposes, a five-year write-off or amortization of all or part of the cost of "defense-certified" facilities. First instituted in World War II and adopted again after the out-break of the Korean war, this device was designed to increase the cash flow of corporations engaged in defense work by permitting them to write off in five years a facility which might have an economic life of as much as twenty or thirty years. The greatly inflated depreciation charge had the effect of reducing taxable net income and, thereby, the impact of income taxes, serving to stimulate activity in the building of facilities regarded as essential for national defense.

---

[4] The charge for depreciation also includes the charge for ordinary obsolescence. There is no set rule for determining the rate of obsolescence.

In the Internal Revenue Act of 1954, Congress extended the principle of fast write-offs—but at a slower rate—to nondefense facilities as well, i.e., to *any* new machinery and buildings. Under the new law, two accelerated methods of depreciation are made available as alternatives to the straight-line method. These are the "declining balance" method and the "sum-of-the-years' digits" method, respectively. The differences among the three methods, none of which require special authorizations like the fast write-off initiated during the war, may be contrasted as follows.

1. *Straight-Line Method.* Under the straight-line method, generally used prior to 1954, the cost of a new machine or building is spread equally over its expected life. For instance, with a machine costing $1,000 and having a life expectancy of 10 years, a company might depreciate the asset at a 10 per cent rate on original cost, or $100 a year.

2. *Declining Balance Method.* Under the declining balance method, the company can deduct up to double the straight-line rate, but on the "undepreciated" balance each year rather than on the original cost. Thus for the $1,000 machine, the second year's deduction would be 20 per cent of the remaining $800, or $160; the third year's deduction would be 20 per cent of the remaining $640, or $128; and so on. This method never permits a 100 per cent depreciation of the asset no matter how long the process is carried on. Therefore, the balance which remains at the end of the asset's economic life is treated, for accounting purposes, as salvage value.

3. *Sum-of-the-Years' Digits.* Under the sum-of-the-years' digits method, which is actually a variation of the declining balance method, a diminishing depreciation ratio is employed which is derived and used in the following manner: (*a*) The years of useful life of the asset are summed, and the resulting figure is the denominator of the ratio. Thus, for the machine example with a life expectancy of 10 years, the total of the digits 1, 2, 3 . . . 10 is 55, and this 55 becomes the required denominator. (*b*) The numerator of the ratio represents, each year, the number of years of life which the asset has, and thus declines by one each year. In the machine example, the numerator would be 10 in the first year, 9 in the second year, and so on down to 1 in the tenth year. (*c*) The depreciation ratio is thus composed of a varying numerator and an unvarying denominator, and this ratio is applied each year to the asset's original cost. In the machine illustration with the original cost at $1,000, the first year's depreciation deduction would be $^{10}/_{55}$ of $1,000, or $181.82; the second year's deduction would be $^{9}/_{55}$ of $1,000, or $163.64; and so on down

to the tenth year for which the depreciation deduction would be $1/55$ of $1,000 or $18.18. By using this method, the depreciation charged declines consistently, and the sum of the depreciation allowances always amounts to exactly the cost of the machine so that the asset is fully depreciated at the end of its economic life.

Table 4–1 and Figure 4–1 reveal some of the important differences between the three depreciation methods.

TABLE 4–1

THREE METHODS OF DEPRECIATION*

Estimated life of asset: 10 years
Original cost of asset: $1,000

YEAR	STRAIGHT-LINE (10%)		DECLINING BALANCE (20%)		SUM-OF-THE-YEARS' DIGITS	
	Annual Charge	Cumulated	Annual Charge	Cumulated	Annual Charge	Cumulated
1	$100	$100	$200	$200	$182	$182
2	100	200	160	360	164	346
3	100	300	128	488	145	491
4	100	400	102	590	127	618
5	100	500	82	672	109	727
6	100	600	66	738	91	818
7	100	700	52	790	73	891
8	100	800	42	832	55	946
9	100	900	34	866	36	982
10	100	1,000	27	893	18	1,000

* Figures are rounded to nearest dollar.

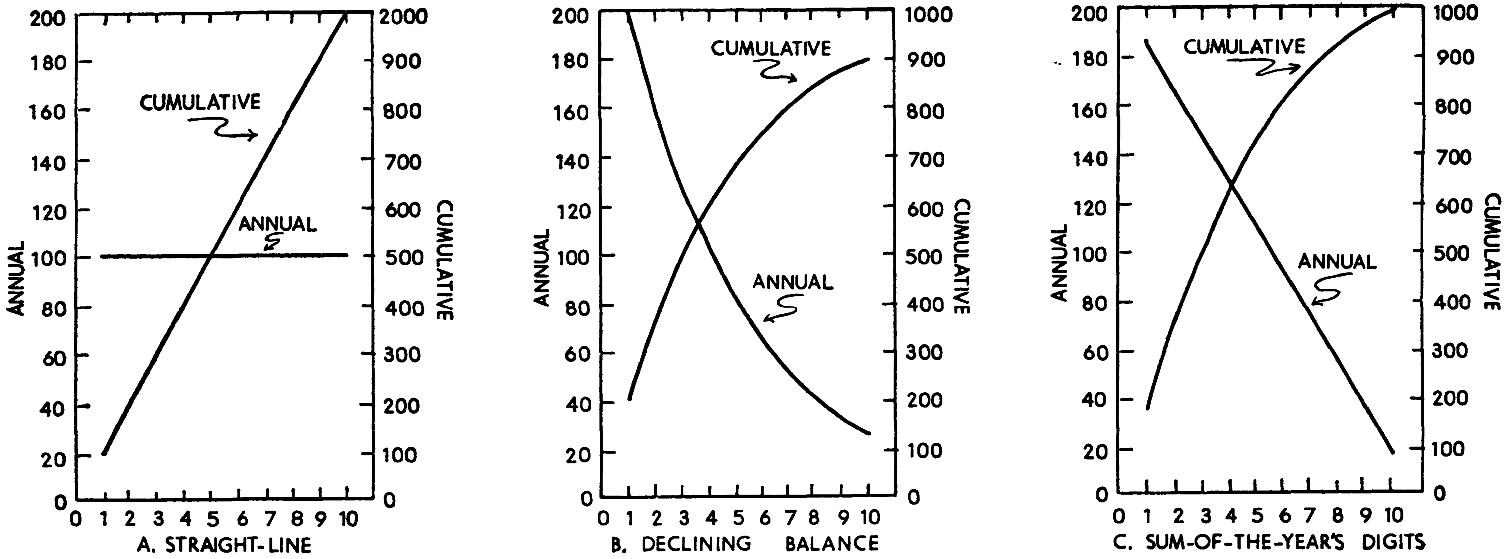
1. Under the straight-line method, half the asset's value is written off at the end of half its calculated life; under each of the two accelerated methods, over two thirds of the asset's value has been written off at the end of half its calculated life. Also, under the accelerated methods there is a rapid falling off in the depreciation charge, which explains why these methods are designed to encourage replacement of facilities sooner than is likely to take place under the straight-line method.

2. Under the declining balance method, the asset cannot be fully depreciated because only a percentage (in this case 20 per cent) of the *undepreciated balance* is written off each year. Further, the depreciation charge is heavier in the first year than under the sum-of-the-digits method, but because the latter leads to a complete write-off at the end of the asset's useful life while the former ends up with an undepreciable balance, the depreciation charge falls off more rapidly under the declining balance method.

3. For very long-lived assets (e.g., 50 years) the accelerated methods do not produce such striking results, for even a doubling of the straight-line annual rate would lead only to a 4 per cent write-off under the declining balance method, and slightly less than that (3.92 per cent) under the sum-of-the-digits method.

FIGURE 4–1

THREE METHODS OF DEPRECIATION



A. STRAIGHT-LINE

B. DECLINING BALANCE

C. SUM-OF-THE-YEAR'S DIGITS

Source: Table 4–1.

4. The first year's depreciation charge is relatively larger under the declining balance method than under the sum-of-the-digits method the shorter the life of the asset, but the difference narrows rapidly as the asset's life increases until, for very long-lived assets, the first year's depreciation is virtually the same under the two methods.

The accelerated depreciation provisions may be applied only to tangible property having a minimum useful life of three years, and became effective as of January 1, 1954. However, there are disadvantages as well as advantages attending the use of accelerated depreciation, and only future developments will reveal whether or not the choice of one of these methods will, in the individual case, prove to have been wise. These considerations relate particularly to tax policy and are discussed below.

Depreciation and Tax Policy. Manifestly, the accelerated methods provide larger depreciation charges in the early years of the asset's life, and correspondingly smaller taxable income and taxes than is the case with straight-line depreciation. If the asset in question is kept in the business for all or most of its useful life, the depreciation charges would fall off rapidly in the later years to levels substantially below those that would prevail under straight-line depreciation. Assuming no change in tax rates or in income before depreciation, taxable income and income taxes would be substantially larger, thereby offsetting the lower taxes of earlier years. All other things being equal, however, the corporation would still have the advantage, under the accelerated methods, of having had the productive use of cash that would otherwise have been paid out in taxes had the straight-line method applied. This cash, in effect available to the company as an interest-free loan from the federal government, could, depending on the useful life of the asset, be employed in the business for a number of years for any of the numerous corporate purposes upon which management might decide, thereby reducing the need for outside financing.

Whether a company's choice of one of the accelerated methods will, in the future, prove to have been wise, depends a great deal on at least two factors, each of which is subject to change. Since the accelerated methods result only in a postponement of taxes rather than a permanent avoidance of them, the wisdom or folly of adopting a given course of action depends on: (1) the income tax rates prevailing at the time the deferred tax has to be paid, and (2) the level of corporate taxable income at that time. Corporate management must, therefore, evaluate the future in terms of these uncertainties when adopting a given depreciation policy. With respect to the course likely to be followed by income tax rates, no single corporate management is in a particularly superior position for predicting their level at any time in the future. However, the international situation and our domestic economic policies both tend to inject an inflationary bias into our economy which, if too rapid an erosion of the purchasing power of the dollar is to be prevented, calls for a continued high level of taxes.[5] On the other hand, the dangers of an onerous tax are well recognized so that at any given time when tax rates are already high, as they are today (late '50's) the probability of their going much higher is rather small—except for temporary measures, in the form of something like an excess profits tax, in unusually critical periods. Conversely, when and if tax rates are low, the possibility of their going substantially higher becomes quite real.

With respect to a given company's future income, no one is in a better position to evaluate this than the company's own management, for business income is, at least to some extent, subject to management's control. If income is expected to increase, then other things being equal, future tax payments will be greater than present ones, and this consideration alone would favor the adoption of straight-line depreciation. But other considerations complicate the picture so that the choice becomes necessarily individualistic, varying from one company situation to another. Among the more important complicating factors are: (1) current versus anticipated future working capital requirements, (2) extent and timing of planned capital expansion programs, and (3) the fact that a present dollar is worth more than a future dollar, i.e., the company's cost of capital.[6]

In light of the foregoing, it would seem that accelerated depreciation offers a clear advantage to young, growing companies with limited access to capital markets and relatively great needs for immediate funds to finance expansion. Even well-established companies with excellent credit will find this an attractive alternative to straight-line depreciation if these companies are engaged in a program of rapid and continuing capital expansion and/or are subject to very rapid plant obsolescence. The latter is particularly true of the chemical industry, where new products rapidly replace existing ones, and in the oil industry, where the race for higher octane gasoline makes obsolete refining facilities unable to produce the upgraded product. Furthermore, as long as a rapid rate of growth is anticipated, accelerated depreciation on new facilities will serve to make up for the rapid decline in the depreciation charges on facilities put in place a few years earlier. But managements of such rapidly growing enterprises will have to realize that as soon as the rate of expansion starts to flatten out, a sudden "burst" of taxable earnings will occur and, if tax rates are still high at the time this happens, the tax bill will be inordinately large. On the other hand, this in itself need be no cause for alarm if it is also realized that the adoption of accelerated depreciation by such expansion-minded managements will have made possible a rate of growth which could not otherwise have been achieved—at least not as rapidly, and certainly not as cheaply.

Because the fiddler must eventually be paid, most of the companies that were granted rapid amortization of defense-certified facilities have attempted to "normalize" the earnings which they report to their stockholders. Thus, instead of reporting earnings after deducting the accelerated charge (as is done on their tax returns to the Bureau of Internal Revenue) these companies compute earnings after "normal" (straight-line) depreciation and deduct, also, a charge in lieu of taxes which did not have to be paid because of the use of accelerated depreciation. This method results in the establishment of a tax "reserve" which recognizes the future (probably larger) tax liability. The adoption of this method has been rather widespread throughout many industries—aluminum, steel, airline, chemical, oil, and public utility—and has been endorsed by the American Institute of Accountants and by the Securities and Exchange Commission. It has been applied at first mostly to those facilities covered by Certificates of Necessity, but is being more widely adopted with respect to accelerated depreciation under the Revenue Code of 1954.

---

[5] This point is discussed in greater detail in the next subsection.

[6] The concept "cost of capital" is discussed in the chapter on capital budgeting. This cost varies from one company to another.

Price-Level Changes and Asset Valuation

In preparing balance sheets and income statements, accountants operate on the "going concern" convention that the business will continue indefinitely. Hence, on the assumption that the company will not sell its fixed assets, it is customary to value these in terms of original cost rather than current market value. Therefore, depreciation charges represent what may be regarded as a proration of historical dollar cost.

While conservative accounting practices prevent market fluctuations from entering into the fixed asset accounts, attempts to recognize what were believed to be "permanent" price-level changes have at times been made. Thus, fairly widespread write-ups of plant were effected in the 1920's, and were as widely reversed in the 1930's. With respect to inventories, however, conservatism has led to the development of the "lower-of-cost-or-market" rule which, interestingly enough, does recognize downside price fluctuations of sufficient amplitude. This special treatment accorded to inventories is perhaps justified because during the production process a certain amount is continually being used up and replaced. If prices were constant and the size of stock always the same, accounting for inventory use would present no particular problem. But when prices fluctuate, inventory replacement at varying cost levels raises the problem of measuring the costs to be applied to the utilized inventory. Accountants have devised two methods of measurement, commonly called first-in, first-out (FIFO) and last-in, first-out (LIFO).[7]

Valuation by FIFO. Under the FIFO method of valuation, the production sequence is viewed as a continuous historical process. The units that are the first to go into the plant as raw materials are also the first to come out of the plant as part of the finished product. Hence, when prices are rising, the goods used up are costed out at the earlier, lower price levels so that the operating statement reflects an inventory profit, and the remaining unused inventory is carried at the more recently prevailing prices. Conversely, when prices are declining, the higher-cost inventory acquisitions are charged against current operations resulting in a narrowing of profit margins, or even a reporting of operating losses, and the remaining inventory is valued at the lower prices at which the material was recently acquired. This was the state of affairs, with respect to inventory accounting, until the advent of World War II. Because it seemed to comply with the actual way in which a business managed its physical inventory, getting rid of its old stocks first and keeping on hand the fresh, most recently acquired materials, the FIFO method was logically correct and almost universally followed. The criticisms directed at FIFO inventory accounting—that it permitted inventory profits and losses (as the result, respectively, of rising and falling prices) to distort the "true" picture of a company's operations—were not taken seriously enough to have any great impact on business accounting practices. Not, that is, until World War II made it quite obvious that the prewar cost structure was rapidly becoming an antiquated relic, and that the price level was reaching a plateau from which there was not likely to be a return.

---

[7] There are other methods such as the average-cost method and the identified-lot method, but these will not be discussed here. FIFO and LIFO are most common

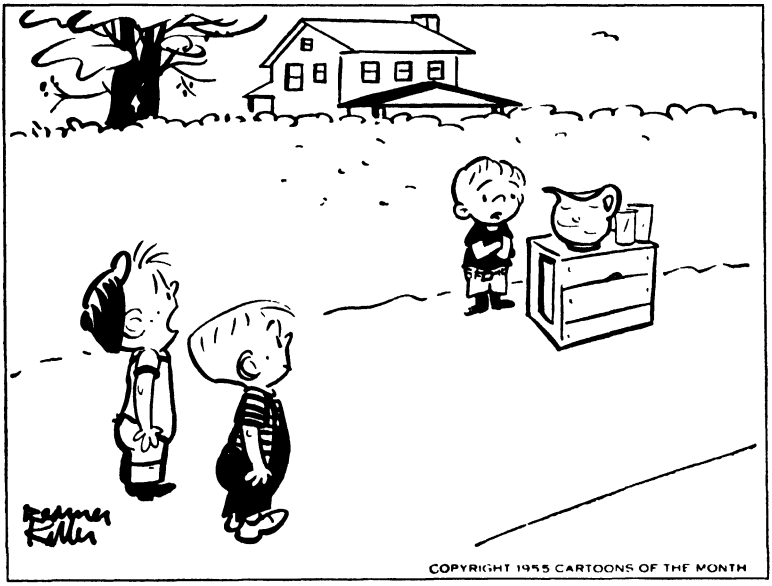
The sharp rise in the general price level that began in 1940 and continued into the early postwar years led to seriously distorted results in corporate financial statements, and raised a considerable amount of discussion among businessmen, accountants, and economists as to the proper treatment of assets. For instance, Joseph E. Pogue, vice president of Chase National Bank, said, "It thus becomes apparent that the changing value of the dollar distorts the income account so that the reported net income ceases to be synonomous with profit."[8] And Eugene Holman, president of Standard Oil of New Jersey, commented, "Our depreciation allowances are based on original cost. Therefore our accounting profit does not give now, as it did before the war, a measure of the funds available for increased capacity and for dividends."[9] And finally, in a testimony before the Presidential steel board regarding the labor-management dispute in the summer of 1949, the noted accounting authority, Professor W. A. Paton, representing the Republic Steel Corporation, said that during periods of continuous price rises, the reported net incomes of corporations tend to be overstated, while depreciation, cost of goods sold, and the book value of stockholders' equity tend to be understated.

Valuation by LIFO. The remarks quoted above reflected the concern of most businessmen with the effect of price rises on inventory accounting and on depreciation allowances. The response with respect to the latter culminated in revision of the revenue laws as discussed earlier, though pressure continues to be exerted for other changes to be considered shortly. With respect to inventory accounting the result was a rather widespread adoption of LIFO, with the underlying reasoning running somewhat as follows. Under the LIFO method, the last units acquired in inventory are the first to enter production. This means that the prices paid for the last units become the costs of the raw materials in current production. From this it follows that if the firm maintains a fairly stable inventory, the cost of raw materials is always close to market value, and only when the inventory is reduced do the earlier purchases of stock enter into the computation. Consequently, in a period of rising prices, the LIFO method yields a higher cost of goods sold since the most recent acquisitions are the first to be costed out in production. The results of these higher costs is to reduce the profit increase in a period of rising prices. Conversely, when prices are falling, the last units acquired are the first to enter production, so costs are thus lower and the profit fall is reduced. As compared to FIFO, therefore, which tends to magnify the profit increase in periods of rising prices and the profit decrease in periods of falling prices, LIFO, it was argued, would act as a restraining influence and stabilizer, by holding back both a profit increase in prosperity and a profit decrease in recession.

---

[8] See Machinery and Allied Products Institute, *Bulletin No. 2138*, January 21, 1949.

[9] *Ibid.* This outlook was somewhat mitigated, however, in *Accounting Research Bulletin No. 33*, and in some published reports of the American Accounting Association.



COPYRIGHT 1955 CARTOONS OF THE MONTH

"He'll fold in a month. No working capital, overextended on inventory, and an assets-to-liability ratio of only one half to one."

Both in business and in academic circles, an almost naive enthusiasm was engendered among those who felt that the cure had at last been discovered which would solve what had been an ever-present problem in inventory valuation. For, although LIFO valuation was an artificial approach contrary to the business practices of maintaining inventories as fresh and new as possible, the accounting function did not necessarily have to be tied to actual business practice. Besides, there was the obvious advantage of ironing out fluctuations in the profit and loss statement to the extent that they stemmed from inventory price movements.

Where naivete did exist, it was among those who failed to realize that only in periods of more or less normal price and inventory changes did LIFO really work as was expected of it, and that under the wrong conditions distortions in the profit and loss statement were much more serious than would likely be caused by the condemned FIFO method. These conditions are: (1) a sharp drop in price, bringing the level below the cost basis of the original inventory established when LIFO was first instituted; and (2) a decline in the physical stock to the point where earlier and very-low-cost inventories are brought into the cost of goods sold. In the first case, the sharply lower price forces a revaluation of the original inventory stock (through the application of the "lower-of-cost-or-market" rule) with concomitant inventory losses to be recognized. In the second case, very strange results become possible. For example, a company may have been on LIFO for a ten-year period during which prices were moving up steadily and operations were proceeding at a pace which permitted stocks to be maintained at desired physical levels. Labor difficulties set in, and a strike is called which forces the company to operate out of inventory for a prolonged period. Soon the inventory which has been carried at prices which prevailed ten years earlier is brought into sales, and huge inventory profits are realized. It is even conceivable, in fact, that these very large profits result in reported earnings far in excess of those realized for the equivalent period before the strike began. For these reasons there has developed evidence in some quarters of a disenchantment with LIFO valuation and a desire to return to what is felt to be the more logical and realistic FIFO approach.

Effect on Depreciation. With respect to the recognition of fixed asset depreciation, price-level changes also work great havoc. The previous discussion of the new methods allowed under the 1954 law abstracted from price changes, for the newer methods, as is true of the older straight-line method, are tied to historical cost. While accelerated depreciation provides a large cash throw-off during the early life of the asset, the sharp drop in depreciation charges in later years results, as we have seen, in sharply higher taxes (other things being equal) thereby offsetting in large measure the lower tax payments in earlier years. Businessmen are still concerned with the fact that as capital expenditures taper off depreciation charges will likewise decline and events will catch up with them in the form of higher taxes. This in itself is no justification for directing criticism at the new law as making inadequate provision for the needs of business. The law does, at least, make it possible to postpone the payment of taxes and in this respect it fulfills a pressing need for all businesses engaged in heavy capital expansion programs, and is particularly beneficial to smaller companies which are going through a period of rapid growth. But there is still legitimate cause for complaint on the part of businessmen who see the forces of inflation consistently eroding the purchasing power of their replacement funds and who recognize that consistent growth of dollars is necessary just to stand still in terms of physical facilities. If we accept the premise that this will continue to be a long-term problem because we have a long-term inflationary bias built into our economy, then no depreciation method will make available sufficient replacement allowances so long as it is tied to historical cost. That the premise of a long-term "built-in" inflation mechanism in our economy seems reasonable is borne out by recent events (price inflation has been going on since 1940) and has sound theoretical support.

As is well known, we are committed to a policy of full employment to be achieved within a framework of free and collective bargaining. These two objectives can be achieved (almost) necessarily at a sacrifice of stable money. A shift to a stable money policy, on the other hand, requires a sacrifice of one or both of the other objectives. Since our national policy has emphasized full employment without undue restrictions being imposed on collective bargaining, we must learn to live with inflation and to deal with it as effectively as possible. The near-term situation is further aggravated by economic and political "parameters" substantially beyond our control. The "cold war" forces a high level of spending for military purposes at a time when demands for other welfare and internal purposes place heavy strains on the federal and local budgets. But not quite as clearly recognized is the fact that severe pressures are building up as a result of persistent shifts taking place in the composition of our population. On the one hand, medical advances have greatly reduced the incidence of death, with particularly telling effects among the very young and the old. Both of these segments contribute nothing, of course, to the supply of goods and services. On the other hand, as people retire from the labor force they are being replaced by others who were born during the "thirties" when the birth rate was abnormally low. The pressures currently exerted on our labor force stem, in large measure, from this low birth rate, and will continue with us for the next several years. This, however, is merely an aggravating, not a fundamental force. When this pressure is finally alleviated inflation will still be with us if our national economic policy has not changed.

It is, therefore, important that our tax policy take cognizance of this "built-in" inflation feature and its effect on replacement funds for plant and equipment. Facilities put in place twenty or twenty-five years ago, replacement of which might now be contemplated, have furnished far from adequate depreciation allowances because of the sharp rise in construction costs that has taken place during the period (such costs have more than tripled). To the extent that allowances are inadequate, the deficiency must be made up from retained earnings or from new financing. In any case it is clear that reported earnings are substantially overstated and that true costs are greater than they appear on the income statement. Such costs can, in the long run, be recovered only by shifting them to consumers in the form of higher prices. At the same time the continuing overstatement of earnings provides labor leaders with motives and arguments for wage increases in excess of what productivity improvements alone might justify. While such demands can hardly be attributed directly to our tax policy, it seems reasonable to believe that the latter is at least a contributing cause.

With a 52 per cent tax on corporate income, every dollar of retained (after-tax) earnings used to supplement a deficient depreciation reserve represents about $2.08 of pretax income. Clearly, the long-run effect of forcing such deficiencies to be made up out of retained earnings must necessarily mean a higher price level than would prevail if depreciation allowances were realistically computed. A realistic approach would require that depreciation be tied to replacement rather than historical cost—a method not yet permitted by the tax laws. Only in this way could the effects of inflation on depreciation allowances be properly offset.

Replacement cost accounting, while not allowed for most assets, is actually being performed when the LIFO method of inventory valuation is employed, and can be reasonably extended to long-term assets as well. However, because of the undesirable aspects of LIFO, a more generalized replacement cost approach would be preferred, to be uniformly applied to both inventories and fixed assets. The essential idea is to arrive at a reported profit figure that reflects the revenues and costs of the present period, not the revenues of the present year and the costs of previous years. For practical purposes under the present circumstances, perhaps the best method of attack is to adjust the data by the application of index numbers. This is readily enough done by making the adjustments with the aid of appropriate indexes, depending on the nature of the asset or account being adjusted.[10]

We will close this section with a rather interesting comment from a challenging work by H. W. Sweeney, *Stabilized Accounting*. Sweeney writes (pp. xi–xii):

> . . . in greater or less degree, ordinary accounting figures give bad advice. They say to expand or contract, buy or sell, hire or fire when sometimes the opposite should be done and when usually the extent of such action should be modified or enhanced. They say that depreciation and costs are such and such when they are more or less. They frequently say that income taxes should be paid when real income indicates they should not be, and *vice*

---

[10] For example, plant assets might be adjusted by a construction index; agricultural implement sales by an agricultural equipment price index; etc. There is a wide area for research in finding the most suitable index for each purpose.

versa. Consequently, business uses a guide that is certainly not wholly reliable when it uses accounting.

PROFIT FORECASTING—CONTROL

There are three approaches to profit forecasting and control that are in common use by business economists: (1) break-even analysis, (2) time-series projections, and (3) correlation analysis. Each of these may be used separately or in combination with others depending on the information known and the purpose of the analysis, and are described below with these considerations in mind.

Break-Even Analysis

A technique of profit planning that came into use a number of years ago and has since gained increasing popularity among accountants, businessmen, and some economists, is that of break-even analysis. This is essentially a graphic device (but equivalent algebraic methods also exist) for integrating costs, revenues, and output of the firm so as to illustrate the probable effects of alternative courses of action upon net profits. The technique contains many variations and applications that are adequately described elsewhere;[11] in the paragraphs below only a few of its essential characteristics are highlighted in order to provide a basic understanding of its nature.

The economic basis of break-even analysis stems from the cost-output and revenue-output functions of price theory illustrated in Figure 4–2. These curves are also familiar to the reader from his elementary education in economics. The diagram shows the total revenue curve *TR*, the total cost curve *TC*, and the corresponding net profit curve *NP*, as these relationships are commonly expressed in the literature of economic theory. They represent the short-run cost and revenue data for a single firm under static conditions, i.e., a fixed plant, no change in technology, or, in general, "a given state of the arts." The total revenue line, determined by price per unit times the number of units sold, is curved concave to the base, indicating that the firm can sell additional units only by charging a lower price per unit on all units sold. (If the firm could sell additional units at the same price, as in pure competition, the *TR* curve would be a straight line.) Total revenue starts at zero output indicating that when there is no output there is no revenue. Inventories are assumed not to exist so that the firm sells all it produces.

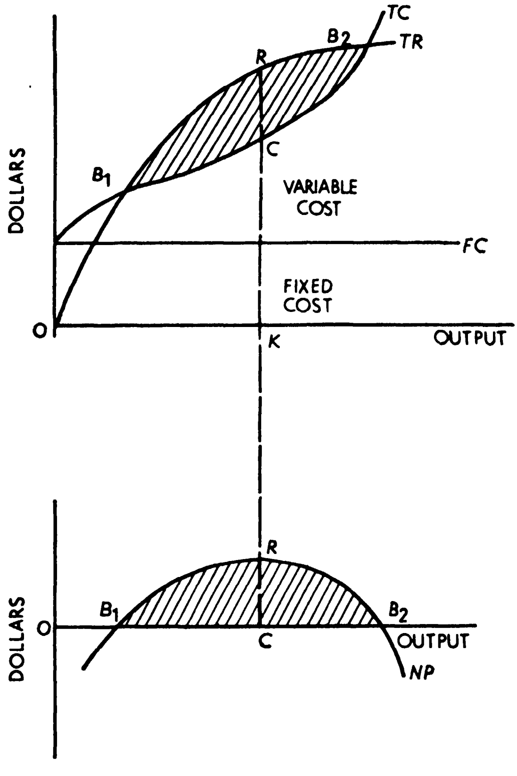
The total cost curve represents the sum of both fixed costs, *FC*, and variable costs. Fixed costs are those costs which do not vary with (are not a function of) output. They include "franchise" payments such as real estate taxes, contractual payments such as rent and interest on capital for the use of specific resources over a fixed time period, and all other constant payments for flow services given off by fixed resources during the production period and irrespective of the level of output. Variable costs are those costs that vary with (are a function of) output. They include all payments made for the flow of services given off by resources in the production period, but which vary according to the level of production. Examples of variable costs are direct labor and raw material expenses. (These concepts are discussed further in Chapter 7.) In Figure 4–2, the variable cost area lies between the *TC* and *FC* curves. In short, total cost equals fixed cost plus variable cost. As in other functional relationships in economics, to be discussed in subsequent chapters, the total cost curve or cost function represents the dependent relationship between cost and output.

---

[11] See the bibliographical note at the end of this chapter.

FIGURE 4–2

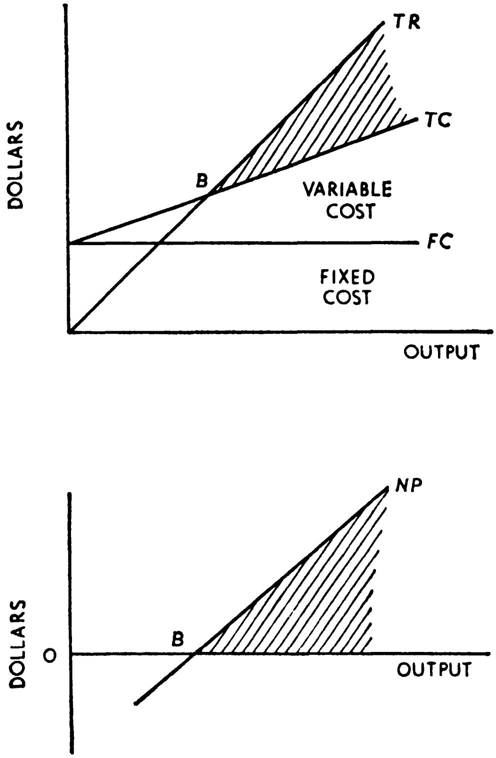
GENERALIZED COST, REVENUE, AND NET PROFIT CURVES



The difference between total revenue and total cost represents net profit, *NP*, and is shown by the shaded area in Figure 4–2. When the net profit (and loss) data are plotted, the result is the *NP* curve. The segments of the *NP* curve below the horizontal axis represent losses or negative profits. Net profit is a maximum where $TR - TC = \text{maximum}$, as at the output OK in upper panel ($= OC$ in lower panel) where the vertical distance RC is greatest. The chart reveals two break-even points, i.e., two levels of output at which the firm's revenues just cover its costs so that net profit is zero. These are the points B_1 and B_2. At these two output levels the firm is receiving only normal profits, since costs are assumed in economics to be determined by the returns to productive inputs in alternative employments, i.e., opportunity costs.

FIGURE 4–3

Break-Even Chart and Net Profit Curve



This is essentially the way in which the cost, revenue, and profit functions, as well as the break-even points, are portrayed in static economic theory. We turn now to what is sometimes said to be a more practical formulation of this construction, the break-even chart.

Break-Even Chart. With a few modifications, the upper panel of Figure 4–2 forms a basis for the construction of the break-even chart shown in the upper panel of Figure 4–3. The nature of these modifications rests mainly on two assumptions.

1. If further units of the product can be sold at the same price, the firm's revenues would now be represented by a linear total revenue (sales)

curve *TR* emanating from the origin. This would apply to a firm in a purely competitive industry, for example, since it is a fundamental assumption of pure competition that no one firm is large enough to influence the market price by offering or withholding its output. but the case is also applicable to many business firms in other competitive situations (e.g., oligopoly) where the product can be sold without a break in the price, at least over wide ranges of output.

2. If further units of productive services can be purchased at the same price per unit, the firm's costs would now be represented by a linear total cost curve *TC* emanating from the intersection of the fixed cost line *FC* with the vertical axis.[12] The assumption of a linear total cost curve may also be quite reasonable within wide ranges of output, as evidenced by various empirical studies of costs (see Chapter 7). The resulting break-even chart in the upper panel of Figure 4–3, reconstructed from its theoretical counterpart on the assumption of linear expense and sales relationships with output thus reveals the profitableness of operations at each output level within the firm's normal production range.[13] Since the sales and cost curves are straight lines, there is only one break-even point which occurs at *B*. The shaded area represents net profit which is also shown by the *NP* curve in the lower panel of the chart.

The break-even chart is a static representation in that it illustrates the relationships between costs and revenues at a given time. The total cost curve of the chart shows what the total expenses would be for any given sales volume according to the *present* budget of expenses. From the chart, management can read off the profit or loss that would result from any output volume. Whether the output volume is measured off on the base of the chart in physical units, or in dollar value of sales, or in per cent of capacity utilized, the same interpretations prevail. And, of course, the output volume (in units, sales, or per cent capacity, depending on the measure used) at which the business breaks even can also be determined from the graph, by dropping a perpendicular from the intersection of the *TR* and *TC* curves to the horizontal axis of the chart.

Contribution Profit. Businessmen do not usually think of profit in the economic sense as total revenue less total cost. Instead, for short-run decisions where a portion of the firm's capital is already a sunk investment and hence immobile, they use a more appropriate profit concept known as *contribution profit,* or the difference between receipts and variable expenses. Thus, if a product sells at $1.00 per unit and the variable expenses are 30 cents, each unit sold covers its variable expenses and the remaining 70 cents is contribution profit, since it contributes to the recovery of fixed expenses and the earning of profit. In economic terms, assuming linear cost and sales relationships with fixed costs, *FC*, imposed as a net addition to variable costs, *VC*, the area of contribution profit, *P*, as distinguished from net profit, *NP*, is shown in Figure 4–4. Contribution profit is thus equal to net profit plus fixed costs ($P = NP + FC$), or total revenue less variable costs ($P = TR - VC$). Also, total revenue is seen to be the sum of contribution profit and variable cost ($TR = P + VC$).

---

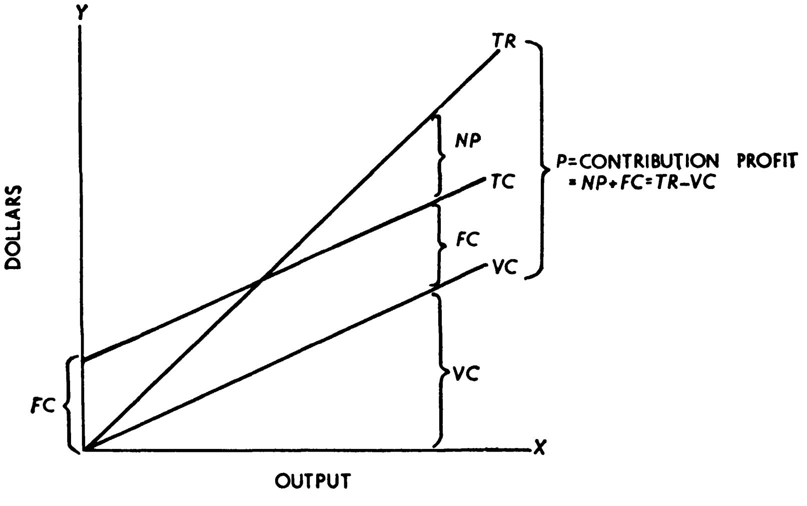
[12] The production function must also be linear.

[13] The assumption of linear relationships, aside from the supporting evidence of recent cost studies discussed in Chapter 7 below, may also be quite reasonable from an analysis standpoint. In mathematics, for example, relationships are frequently linearized to reduce their complexity.

As Figure 4–4 stands, it conveys all of the information commonly used by break-even analysts in profit planning and control. The original data for the construction of the chart are frequently obtained directly from the published profit and loss statement of the firm. Sometimes a single statement is used and the lines are extrapolated backwards on the assumption that linear relationships prevail. Sometimes several statements are employed representing different output levels; the points are plotted as a scatter diagram and the revenue and cost curves are then sketched in as freehand regression lines. On the horizontal axis, any measure of output, such as physical units, per cent of capacity, or dollar sales, may be employed. When the company's income statement is the source of the data, sales are usually the measure of output since the other indicators are not ordinarily given. If a unique functional relationship is known between dollar values on the vertical axis and the measure of output on the horizontal, such that for each value of *X* there is a corresponding value for *Y*, the scales if equated enable the total revenue (or sales) line to be drawn in immediately as a diagonal of the chart at a 45-degree angle to the base. In any event, when used for profit planning, the chart shows: (1) the output required to net a given revenue, (2) the revenue to be expected from a

FIGURE 4–4

CONTRIBUTION PROFIT



given output, (3) the sales volume required to break even, and (4) variations of these concepts in terms of net profit and contribution profit. For levels of output beyond those shown by the diagram, the revenue and cost curves are generally projected on the assumption that the underlying relationships remain unchanged up until the level of full capacity, however defined. This, essentially, is the basic structure and use of the break-even chart. There are also several variations of other charts that can be constructed from the basic diagram in Figure 4–4, and a number of simple algebraic formulas that can be derived as a substitute for graphic methods in solving numerous problems involving decisions as to cost and sales changes. These are beyond the scope of this book, however, and the interested reader can refer to the bibliographical note at the end of this chapter for sources covering these aspects.

Evaluation. Break-even analysis is a general method of profit forecasting and control, based on the assumption that there is a unique functional relationship between the profits of a firm and its level of output. In symbols, a relation of this type may be written for conceptual purposes as $P = f(O)$, which is read, "profit is a function of output," and means that for each level of output there is a corresponding level of profit. Output, as stated earlier, may be measured in terms of physical units, dollar value of sales, per cent of plant capacity, or any other relevant index. Profit, on the other hand, is a more explicit notion in break-even analysis and usually represents the difference between receipts and expenses for the period under study. As an analytical technique, break-even methods have as their chief advantages simplicity, ease of comprehension by management, and relative inexpensiveness compared to other research methods. Most, and sometimes all, of the data required are taken directly from the published income statements of the firm, and hence break-even analyses can be conducted on a monthly, quarterly, and annual basis.

The static profit-output function, $P = f(O)$, which sums up succinctly the notion of break-even analysis, contains certain implications which are the basis for most of the criticisms leveled against the method. As the above equation stands, it states that profit, the dependent variable, depends on output, the independent variable, and hence, given the level of output, the corresponding level of profit could be determined, provided the mathematical equation of the relationship were known.[14] Realistically, however, profit is dependent on a great many factors other than output which the break-even analysis fails to recognize because of the oversimplified construction of its two essential components: the cost function and the revenue function, the difference between which establishes the profit function. With dynamic forces continually at work to shift and modify the underlying elements determining costs and revenues, any attempt to represent these relations in the form of static functions is immediately suspect.

---

[14] That is, the expression $P = f(O)$ can be thought of as stating the functional relationship conceptually. Mathematically, it may be expressed by any one of numerous equations of which one would be, for a linear relationship, $P = a + bO$. When the constants a and b are established from experimental data, the equation can then be used for prediction.

On the cost side, the chief difficulty is this: by assuming a constancy in the state of the arts, i.e., technology, plant scale and depth, efficiency, etc., traditional break-even methods cannot solve the problems of profit forecasting with much precision. A substantial improvement would be made if there were a careful selection of the enterprise and sample period, and if careful adjustments could be made in order to account for changes in factor prices, product mix, cost-output relations, and similar variations that are influential in their effect on profit, but the increased expense and know-how necessary to accomplish this in a break-even analysis overcomes the practical advantages of the method, namely its inexpensiveness and ease of comprehension by management.

On the revenue side, the use of a static revenue function assumes a constancy of sales mixture, selling prices, and proportion of total output allocated to each distribution channel (i.e., channel ratio), at the very least. But even granted that management can control reasonably well the second and third of these factors, changes in sales mixture are due largely to the whims of consumers. Such changes (as well as changes in the channel ratio), where the contribution profit between products or product classes differs and the changes are not closely correlated with output, may seriously distort the static sales or revenue line and hence the profit forecast. In short, the break-even analysis applied to profit forecasting assumes a continuation of the same relative sales and expense patterns, and hence takes no account of uncertainty influences as manifested by probable changes in revenues and costs as business conditions change.

Thus, the assumption that profit is a *simple relation* with output alone, written $P = f(O)$, as presupposed in break-even analysis is an oversimplification of the facts. Profit depends on output, to be sure, but it also is affected by production processes, selling effort, the composition of demand, and a multiplicity of other factors both internal and external to the firm. A more general statement closer to the facts would be to express profit as a *multiple relation*, namely $P = f(A,B,C\ .\ .\ .)$, where each of the letters inside the parentheses represents specified factors affecting profits such as those listed above, and the dots denote other profit determinants which have not been specified. In subsequent chapters we shall have occasion to make frequent distinctions between simple and multiple relations of various kinds, since a large part of managerial economics is concerned with the measurement of functional relationships (e.g., demand, production, costs, etc.) similar in principle to the type discussed thus far.

Do the limitations stated above mean that the break-even type of analysis is useless? For those firms that experience rapid changes in their main cost components, in their sales mixture, in their advertising and promotional policies, and in their technology and product design, the answer is probably yes. Although methods do exist for building greater flexibility into break-even analysis, such procedures quickly offset its chief advantages of simplicity and inexpensiveness, and these are features that management will not readily sacrifice. Perhaps the best that can be said is that the break-even chart serves its maximum usefulness when used as a supplement to other forecasting techniques (e.g., econometric methods) illustrated in later chapters.

Time-Series Projections

Time-series analysis in the form of income statement projections is another method commonly used in profit planning. Either of two procedures are usually employed: (1) Sales and cost figures are taken directly from past income statements, time series are established, and growth trends are passed through the data by conventional statistical procedures; the profit forecast is then the residual of these sales and cost projections. (2) Instead of forecasting sales and costs first and then taking the difference as the profit forecast, the alternative is to project the past-profit figures directly. Either of the two methods involve the identical statistical procedures, and may also be applied to every item on the profit and loss statement, thereby arriving at a projected-income statement for any period in the future. In the same manner, cyclical and seasonal variations can be measured and the appropriate projected income statements can be built up to show these factors as well. In short, the measurement techniques employ the methods traditionally described in all elementary statistics textbooks.

The use of time-series analysis as a general approach to forecasting was evaluated in Chapter 2 dealing with forecasting methods. Some brief comments, however, can be made at this point. First, the forward extrapolation of secular trend is essentially a projection of past or historical relationships, and hence assumes that the future profits will be affected by the same relative relationships between sales and costs as existed previously. Consequently, it takes no account of changing technology, efficiency, plant scale and depth, etc., as factors affecting costs, nor of changes in product mix, prices, distribution channel ratios, etc., as factors affecting sales. It is essentially a static method and in this respect has shortcomings similar to those of break-even charts. Second, there are the statistical problems themselves such as: (*a*) the assumed relationships between the elements of the time series, particularly as to whether they are additive (i.e., $O = T + S + C + I$),[15] multiplicative (i.e., $O = TSCI$), or whether they stand in some other relationship to each other; (*b*) whether the traditional residual method should be used in isolating the cycle or whether another procedure would be more appropriate; and (*c*) whether the trend should be represented by a straight line or a curvilinear relationship, and the correct choice of curve if the latter is chosen. These considerations as well as others were taken up in greater detail in earlier chapters and need not be dwelt upon here.

---

[15] The equation in question simply means that the original data are the sum of the trend, seasonal, cyclical, and irregular factors.

Correlation Analysis

A third method commonly employed in profit forecasting is correlation analysis. This procedure was discussed in Chapter 3 and can be treated briefly here as related to profit prediction. Essentially, the goal is to discover a functional relationship between the company's profits and one or more indicators of national economic change such as the Federal Reserve Board's Index of Industrial Production, disposable income, bank debits, etc. Frequently, "time" is used as one of the independent variables. The underlying assumption in the use of this method for profit prediction is that the well-being of the firm as measured by its profits is directly determined by business conditions in the total economy; the company, in other words, is a product of its environment. Profits are thus treated as a dependent variable and the relevant measures of national economic change as independent variables.

In practice, this approach to profit forecasting is greatly enhanced when some logical lead-lag relationship can be found between the company's profits and one or more of the external variables. The American Radiator & Standard Sanitary Corp., for example, utilizes the fact that there is approximately a four-months lag between its own sales and the Dodge index of residential contracts awarded, thereby facilitating its profit forecast on plumbing and heating supplies. Where logical lead-lag relationships cannot be found, however, the independent variables must themselves be forecast before a prediction of profits can be made. In that event, the accuracy of the profit forecast will depend directly upon: (1) the accuracy of the forecast made for the independent variables, and (2) the extent to which these external variables are truly related to the company's profits. The first condition can be partially hedged against by selecting indicators that are frequently forecast by various governmental and private agencies—variables such as GNP, disposable income, the FRB index, etc.—so that the various predictions can be crosschecked, weighed, and evaluated. The second condition requires more of a subjective judgment supported by economic reasoning as to which variables will most closely affect the company's present and future earnings. Discovering the relevant data and choosing the appropriate period for the analysis (usually the most recent single business cycle with the data expressed quarterly or monthly) are probably the most difficult aspects of this type of correlation analysis.

NATURE AND DYNAMICS OF PROFITS

Profits Planned and Unplanned

From a managerial standpoint, a central notion in the concept of uncertainty is that anticipations of the future are framed in terms of a range of possible outcomes distributed in some way around a single most-expected outcome. As applied to incomes this means that there will result a deviation of actual returns from planned returns. This follows from the fact that not all future revenues and costs can be known, so that not all resources in the production process can be engaged on a contractual basis. Since the payments to some resource owners are contractual, it follows that other resource owners will receive residual returns. By the very nature of the typical production process, these residual returns will include unanticipated positive or negative components. Profits, it should be made clear, are a mixture of both anticipated and unanticipated residuals. The latter component is a true surplus (positive or negative) and as such is a part of economic profit in the fullest sense. The former component, however, is not necessarily a surplus, and whether it is or not depends on the particular case in question. In the typical well-managed enterprise, the planning process is directed toward planning for profits, and tight financial control and proper capital budgeting require that profits be planned. To the extent that the planned or anticipated profits are just large enough to warrant bringing together the necessary resources for carrying the plan through to fruition (the project would be scrapped if the estimated returns were expected to be less than this minimum)—to that extent they would more properly constitute a functional or compensatory return (long-run cost) and as such would not qualify as "true" surplus.

However, projects are frequently undertaken with the expectation of securing something more than the planned required profits. This extra may, in contrast, be thought of as a planned surplus profit—a return in excess of that necessary to bring the project into being and is, therefore, a "true" profit. In addition, there exist unplanned positive or negative residuals which will increase or decrease the planned surplus. In short, profits may be either anticipated or unanticipated, required or surplus, but in any case they are always uncertain, and will always be so as long as the future cannot be forecast with known error. Were it possible to forecast with known error, uncertainties would then be mere risks and could be shifted to appropriate risk bearers (insured) or be assumed by the entrepreneur in a risk-bearing capacity (self-insured). These elements would be of the nature of costs (normal profits) and the situation would then be similar to that of the stationary state.

Profit Over Time

By its very nature, planning involves the future, and the future introduces the element of uncertainty. In static analysis, profit maximization extends only to the time interval in which business transactions are completed. Thus, if management starts production in "year 1" with the aim of selling in "year 2," its forecasts extend only as far as "year 2." In terms of Chapter 1, management formulates plans in the current time period,

t_1, in anticipation (forecasts) of events that will take place in future time periods t_2, t_3, etc. If the events which are expected to occur within this time period are held with certainty, plans can be formulated in a static vein and no further decisions beyond the initial one of establishing operations would be necessary. Profit maximization would then reduce itself to the problem of making the initial decisions necessary for arriving at maximum profits over the given (planned) time period.

But even in such a situation, a forecast of a $100 profit in some future time period is not the equivalent of a $100 profit in the current one. This is due to the fact that the interest obtainable on a perfectly certain investment, i.e., government bonds, is greater than zero so that there exists a time preference which favors the present as against the future. Thus, if a government bond maturing in one year is available at a yield of 3 per cent, a profit of $100 to be made available one year from today from a production process involving no uncertainty would have a present value of $100/1.03 (= $100/103 per cent) or approximately $97.09. [Conversely, at 3 per cent, the value next year of $97.09 today is 103 per cent of that amount, or $97.09 (1.03) = $100.] Similarly, if a two-year government bond were available at a 3 per cent yield, then $100 obtainable two years from today from a production process with no uncertainty would have a present value of $\$100/(1.03)^2$ or approximately $94.26. If this same process were to produce a stream of $100 to be available at the end of one year and another $100 to be available at the end of the second year, the present value of these two payments would be the sum of their separate present values, or $191.35.

In the real (dynamic) world, however, uncertainty is an ever-present element in the economic environment. Related to the above discussion this means that future receipts must be discounted (capitalized) at a rate in excess of that available on an investment of perfect certainty, with the result that the present value of such future receipts are accordingly lowered. We may express this concept in general terms, namely:[16]

$$I = \frac{R}{(1 + r)^n} \qquad (1)$$

where I represents an investment which will produce a revenue,[17] R, available at the end of n periods in the future, and discounted at a rate represented by r per period (expressed in decimal form). Where a stream of income is expected over a period of years, the equation can be expanded as below, where R_1, R_2, R_3, and R_n represent flows of cash earnings in the first, second, third, and nth years:

---

[16] The symbols used in the equations of this section will appear again in the capital planning chapters where these matters come up for more detailed discussion.

[17] Called "cash earnings" because it includes noncash charges such as depreciation and depletion. In other words, the noncash charges are added back to net earnings giving what is known as "cash earnings."

$$I = \frac{R_1}{1 + r} + \frac{R_2}{(1 + r)^2} + \frac{R_3}{(1 + r)^3} + \ldots + \frac{R_n}{(1 + r)^n} \qquad (2)$$

Just what the profit-maximizing firm seeks to maximize will be discussed in the following section, but at this point we will simply say that it is desirable to maximize the stream of future income. In static analysis, where the current period's profits are simply extended repetitively into the "future," the problem of maximizing an earnings stream reduces simply to the problem of maximizing R. But in a dynamic situation, where fluctuating values for R are projected, the problem of maximization is greatly complicated. However, as a practical approach to the problem, it is often useful to estimate a uniform (average) annual profit and to project this into the future until a change in conditions (such as installation of a new plant) calls for a new projection. Where the flow of expected uniform annual profit, U, is for an indefinite period of time, the following equation defines the value, I, of this income sequence into perpetuity:

$$I = \frac{U}{r} \qquad (3)$$

where r is the "appropriate" capitalization rate.[18] This approach is employed universally in the valuation of stocks by investors who multiply estimated earnings by a factor known as the price-earnings multiple to arrive at an estimated value for the stock in question. The price-earnings multiplier is simply the reciprocal of r, the capitalization rate. It is a technique also commonly employed in the real estate industry, whereby the estimated annual rentals are multiplied by some figure (the reciprocal of the capitalization rate) to arrive at an estimate of the real estate value.

A NOTE ON THE THEORY OF PROFIT MAXIMIZATION[19]

We now have the necessary tools for considering just what the entrepreneur should try to maximize when he sets out to maximize "profits." Unfortunately, the literature has not been consistent in this matter.

Statics

In the static theory of production there is no capital or investment decision making to be done, because the investment problem has been assumed away. In this timeless theory of production, the entrepreneur simply seeks to maximize profits in the sense of total revenues minus total costs. In terms of the previous discussion, as was indicated there, the "earnings stream" from a given investment is simply a repetition of net cash earnings, *R*, produced in each static period into the "future," and it is *R* that the entrepreneur seeks to maximize. This is accomplished by: (1) determining the least-cost combination of production factors (all inputs including plant) for each given output,[20] the least-cost combination being defined by the condition that the relative marginal value products of the factors equal their relative marginal costs (this statement is general enough to cover both perfect and imperfect competition); and (2) moving along the expansion path[21] of least-cost combinations, determined from above, to that scale of output which is optimum as defined by the condition that long-run marginal cost equals long-run marginal revenue. On these conclusions of the static analysis there is uniformity. But not so when "time" is introduced.

---

[18] This is discussed in the next section, and is treated in much greater detail in the last chapter.

[19] The student with no more than a rather elementary background in economic theory should not be chagrined to find this section somewhat difficult to grasp. While it can be skipped without loss of over-all meaning or continuity, all students are advised to read this section at least for general "flavor" if not complete understanding.

Dynamics

The introduction of time as an independent variable in the production function also introduces the problem of capital or investment. But because "time" cannot be treated as just another factor of production the inputs of which can be varied in combination with other inputs to produce constant product curves or levels, the optimum solution cannot be determined in a manner paralleling the static analysis. That is, the problem cannot be approached in terms of minimizing the cost of output levels because an output at one point in the future is different from the same output at another future date.[22]

Thus, because the dynamic analysis cannot satisfactorily be squeezed into the static mold, it becomes necessary to strike out on a fresh path. The most fruitful one involves the capitalization technique discussed in the preceding section. In this approach, as we have seen, an investment is conceived as a revenue-producing process from which flows a stream of cash earnings into the future. The profitability of the process is determined by relating the revenue stream to the cost of the investment, *I*, as shown above in equation (2), and the rate of return for the process is defined as that rate which makes the equation true. In the preceding section we referred to this as the capitalization rate. Now we see that it is also the rate of return on the investment.

---

[20] This is the locus of points at which, for two-factor inputs, a constant product curve touches a constant cost curve. (See also Chapter 6.)

[21] *Ibid.*

[22] Adding dollars or values of output at different dates is similar to adding horses and apples. The results would be equally meaningless. Dollars at different dates are additive only when interest is zero. This requires two conditions: (1) zero time preference, and (2) perfect certainty. Otherwise dollars at different dates must be converted into dollars of the same date (either past, present, or future) in order to add them, and the conversion is accomplished by means of the appropriate capitalization or interest rate.

We now ask: what does the entrepreneur wish to maximize? And the answer, as provided in the economic literature, is not always the same. (For a detailed discussion, employing extensive mathematical tools, the reader is referred to the footnoted references. The conclusion reached there is not, however, the same as ours).[23] Here we will point out that the two most widely suggested criteria are: (1) maximizing the rate of return on the investment (either the total permanent invested capital or, particularly, the entrepreneur's own capital); and (2) maximizing the difference between the present value of the revenue stream and the present value of cost outlays.

In our chapters on capital planning, the argument is developed that the firm can increase its profits by implementing investment projects so long as the rate of return on these projects is greater than the firm's "cost of capital." The profit-maximizing firm which pursues this investment plan will thereby satisfy, simultaneously, both of the criteria indicated above. The reason for the controversy in the literature stems from an inconsistent conception of what a firm's cost of capital really is, so that "arbitrary" capitalization rates are employed under the guise of a so-called "going" or "market" rate of interest. In short, the firm which institutes profit-maximizing policies will apply its cost of capital (defined precisely in the last chapter of this book) as the appropriate capitalization rate and, having achieved a maximizing position, will find that: (1) the rate of return on the last incremental investment equals the firm's cost of capital; (2) the stream of cash earnings produced, when capitalized at a rate equal to the cost of capital, gives a maximum present value in excess of cost of invested capital; (3) the stream of cash earnings produced constitutes a maximum rate of return on invested capital; (4) the stream of cash earnings produced constitutes a maximum rate of return on the entrepreneur's own capital.

The above comments will seem much more meaningful if they are read again after the last two chapters have been studied. At this point it is useful to treat two other aspects of profit seeking and profit maximizing—concepts which will recur on and off throughout this book. This is the distinction that must be made between "marginal cost" versus "incremental cost" and "marginal revenue" versus "incremental revenue."

Marginal Profit

The marginal profit (*MP*) concept is fundamentally of short-run usefulness in that it is applicable to a static economic situation. Simply, it involves a change in profit stemming from variations in output in adjusting to the optimum position under a given state of the arts, given price parameters, and given scale and depth of plant (all of which are often subsumed under the broad use of the term "given tastes"). The criteria for decision making are thus confined to changes in profit as they result from changes in output within the rigidified pattern of a given set of relationships existing between the internal company factors such as plant scale on the one hand and the external factors of product demand on the other. Given these conditions the firm will adjust output to the point where $MP = O$ (i.e., where net profit or NP = maximum).

---

[23] Friedrich A. Lutz, "The Criterion of Maximum Profits in the Theory of Investment," *Quarterly Journal of Economics* (November, 1945), p. 56. See also, Friedrich and Vera Lutz, *The Theory of Investment of the Firm* (1951), chap. ii.

When it is stated that a profit-seeking organization will maximize profits by adjusting its operations as described above, it is implied that the stream of future profits thus produced would, under the given conditions, have a maximum present value when discounted at whatever appropriate rate might apply to the process in question. Stating the matter in this way involves nothing more than dressing up a static concept by employing terms which have a dynamic connotation. But the situation is still essentially static for we are merely projecting the present state of affairs into the future and then simply discounting the future income stream thus "expected." Yet, even though the precise adjustments implied above can never take place in actual business practice, because in the real world the adjustment process would have to go on continually in the attempt to meet and contend with the ever-changing conditions of a dynamic society, the analysis is still conceptually useful.

Incremental Profit

When, on the other hand, significant variations in technology, product mix, sales promotion and distribution, plant scale, etc. are permitted, the analysis does take on certain dynamic aspects, though not necessarily in the strict formalized sense. Greater "realism" is introduced, and with it such concepts as incremental cost and incremental profit become meaningful. Thus, where optimum operations always require that adjustments be made to cause *marginal* profit to be zero, the profit-maximizing firm will never be satisfied with zero *incremental* profit. The latter may perhaps best be defined as the change (increase) in average annual profit resulting from the implementation of managerial decision with respect to any one of a number of possible aspects of the firm's operations: diversification into a new product line, differentiation of the existing product, change or broadening of the distribution channels, increased sales effort, new cost-cutting devices, and so on. After instituting the new activity, the firm will adjust to the new conditions so that marginal profit is again zero, but the level of profit will have changed as a result of the new activity. The difference between the new profit level and the previous one is the incremental profit.

By way of illustration, a firm is estimated to earn an average annual level of, say, $2 million. A new "project" or "activity" is instituted which is estimated to raise average annual earnings to $2.1 million after all necessary adjustments in operations have been made. The (annual) incremental profit is, therefore, $100,000 (note that marginal profit will, after all necessary adjustments, again be zero).

The "incremental" concept may be further extended. Thus, the incremental revenue would be the average annual added revenue resulting from the implementation of the new activity; the incremental cost is the average annual added cost incurred to produce the incremental revenue. Going one step further, if we apply this analysis to a corporation which, in the typical case, operates under a charter in perpetuity and, therefore, produces a perpetual stream of income, we can arrive at an incremental net worth by capitalizing the expected incremental profit, ΔU, at a rate equal to the firm's cost of capital. If, as in the above case, the appropriate capitalization rate for the process in question where 12½ per cent, the incremental net worth (as measured by market value rather than book value) would be in terms of equation (3) above, $I = \frac{\Delta U}{r}$ or $100,000/.125 = $800,000.

When applied to such flows as profits, outlays, and revenues it is best to define and use the corresponding incremental concepts in terms of the time periods ordinarily employed (per year, month, week, or quarter). This is more meaningful and more manageable than to define incremental profit, for example, as the total profit flowing to the enterprise over the life of the activity. There are several reasons for avoiding the latter formulation, not the least important of which is that profits which appear at different points in time cannot be added directly together. They must first be converted into dollars of the same time period, via the capitalization rate, which was of course done above in the illustration of incremental net worth.

In conclusion, profit maximization requires that marginal profit *always* be equated to zero. But incremental profit must always be positive, and whether or not the firm will be enticed into going after this increment will depend on whether the economic effort required will be justified by the added rewards in light of the uncertainties affecting the production process in question. This, in effect, is the basic problem to be solved whenever a firm considers the construction of a new plant, instituting a new product, or entering a new market.

PROFIT POLICIES

We have seen, then, that in economic theory we employ the profit-maximization principle as a basic premise. This is so because a competititive economy in which profit maximization is the guiding rule for business decision makers will achieve optimum efficiency in production, and optimum allocation of its resources. In terms of economic welfare these are desirable economic goals of any society in which material aims ride high. Since profit maximization is thus a basic premise on which to build the economic policy of the firm, it shall be employed throughout this book, recognizing at the same time that departures from it may frequently be taken. In this section, in fact, we consider in some detail the possibilities of such departures, and the alternatives necessitated by them.

Difficulty of Recognizing Aberrations

It is extremely difficult to assert unequivocally that business firms do not strive to maximize profits. For one thing, it is necessary to segregate short-run from long-run policies, and it is clear that many policies which may reduce short-run profits are designed to establish a better long-run situation. In this regard we may point to such programs as (1) aggressive research for new products, (2) costly development of new markets, and (3) fringe benefits to employees aimed at developing long-run loyalties. Another complicating factor which makes an unequivocal answer impossible is that each company approaches the profit problem differently and what to one enterprise may seem a wise policy is deemed unnecessary and, perhaps, even folly to another. This is true particularly in the field of employee relations: pensions, health and accident insurance, and even coffee breaks. One considers such benefits as important in raising labor productivity; another considers them at best a necessary evil.

Nevertheless, it does seem clear that the maximizing principle cannot be accepted without qualification. Studies conducted in recent years have pointed to the conclusion that profit maximization is frequently not the ultimate goal of management. To be sure, in industries characterized by pure competition, the horizontal demand curve confronting the seller leaves little if any room for price discretion, and maximum profit becomes synonymous with normal profit in the long run. But in most of industry where oligopoly market structures prevail in one form or another, the drive to maximize profits is often modified and compromised with other objectives that may reduce earning power. They may be referred to as limitational factors which, in general, are believed to reduce profits below the level which would have prevailed had maximizing motives been the sole driving force. Yet, in almost every case, it is possible to argue that these factors are only superficially limitational, and that the goal of attaining maximum long-run profit levels requires that they be blended into the operating picture as part of the total pattern of internal and external forces which management must reckon with in its drive for maximum profits. This must be kept in mind by the reader in his evaluation of the discussion which follows.

Profit Limiting Factors

A number of causes have been suggested as factors responsible for limiting management's drive for profit maximization. These may be classified into two distinct but related groups: (1) those that are largely internal to the firm's operations, are often indirect in their effect on curbing profit, and may go unrecognized by management as profit-limiting factors; (2) those that are usually external in nature, are known and recognized by management, and may be serious enough to warrant the establishment of plans that will specifically provide for nonmaximum profit. Although this classification may allow for some overlapping, it appears to be fairly reasonable for outlining some of the more common limitational factors.

In the first group are the actions of management that serve in an indirect manner to dampen profits.

1. *Desire for Company Prestige.* Some managers place (excessive) emphasis on establishing the firm as a leader in its industry, even at the expense of lower profits, on the supposition that the company's sales growth is its best measure of success. This occurs, for example, when management devotes a disproportionate amount of effort to broadening the product line by introducing new products, without sufficient attention paid to costs, in order to build a reputation of being largest in the field.

2. *Resistance to Change.* Managers often have a fear or reluctance to make a decision when the expected outcome is other than a near certainty. Rather than chance the possibility of penalty (such as loss of status or even job) in the event of an unfavorable outcome, they prefer to operate on an assumption of "nothing ventured, nothing lost." Preserving the status quo for the security that it offers is to them more important than the "risk" of sacrifice that is attendant with progress.

3. *Excessive Desire for Liquidity.* Where increased profits depend on entering new areas of production, this often requires an increased investment in fixed assets and hence a reduction in liquidity. Some industrial companies have balance sheets that look more like bank statements and, of course, reflect extreme pessimism over the business outlook. Such companies have consistently lost ground to their more aggressive competitors. For many managers, "sound" financial conditions are more important than maximum profits.

In the second group are those reasons why a firm consciously and purposely avoids maximizing (short-run) profits, although the execution of these policies may be argued to be best in the long run.

1. *Discourage Competitive Entry.* If profits could be large due to higher prices rather than lower costs and superior efficiency, or if the company has a weak monopoly position in the industry, management may prefer lower profits in order to discourage potential competitors from entering the industry. In this case a long-run price policy that is in line with the rest of the industry will be more advantageous to the firm than one which exploits current market conditions for immediate profit.

2. *Discourage Antitrust Investigation.* Profits have been one of a number of criteria sometimes employed as evidence of monopolistic market control. This can seem somewhat of a paradox when contrasted

with the previous consideration. On the one hand, management may maintain lower profits in order to exclude competitors and thereby strengthen its monopoly control. Yet the antitrusters consider high profits, not low profits, as one of several indexes of monopoly power.

3. *Restrain Union Demands.* Reducing the possibility of having to pay higher wages is another factor prompting management to restrain profits. This is particularly applicable in industries with strong labor unions. As long as the economy is prosperous and profits are rising, unions can more easily demand a higher wage rate without inflicting damage to the firm. But in a recession when prices are falling faster than wages, the profit margin is squeezed at both ends. Those companies that curbed wage increases in the beginning would then have a better opportunity to cope with changing market conditions.

4. *Maintain Consumer Goodwill.* Management may choose to limit profits in order to preserve good customer relations. Consumers frequently have their own ideas as to what constitutes a "fair" price, whether such ideas are based on "what used to be in the old days," or whether they are the results of "comparison shopping." Many an automobile dealer, for example, can testify as to the customers he lost in the late 1940's when conditions were back to normal, because of an unwise short-run sales policy he followed immediately after the war.

Profit Standards

As a matter of reality, businessmen do not always profess a desire to maximize profit. "The belief that the top managements of large corporations have a single-minded devotion to profits is one of the great myths of modern American capitalism. . . . A company's top executives are likely to put a lot of things ahead of profits, and what's more the pursuit of these other objectives may seriously impair the company's earning power."[24] There is even an occasional antagonism on the part of some managers to a profit that seems "excessive," though more often a strong resistance to a rate of earnings that is less than something which they would call "fair."[25]

It does seem clear that the profit-maximization principle is a questionable premise. When management itself admits to nonprofit motives in its decision making, and when these motives rule to an extent which causes others to issue warnings about the harm likely to follow from "going off the profit standard,"[26] it would seem foolish for anyone to deny that such departures do take place. There are, of course, times when the drive for profits must give way, at least somewhat, to other factors—national defense in time of war, and responsibility to the community. But these considerations should result in no more than temporary or fairly minor departures from what can be the only meaningful measure of corporate success and managerial effectiveness. Since the decision-making role of management must be cast in a profit-seeking framework, it is desirable to employ some kind of profit yardstick as a measure of what constitutes acceptable performance for a given enterprise from management's own point of view as well as the stockholder's. While "maximum profit" is suitable, to the extent that it is realistic, as an economic principle, it is unwieldy as a standard because, as pointed out earlier, it is possible to support every decision and action as one aimed at maximizing long-run profits even though it will obviously result in reducing current or short-run profits. Besides, how can we hope to recognize "maximum" profits even if we were to see them?

---

[24] P. Stryker and the editors of *Fortune*, *A Guide to Modern Management Methods* (1954), p. 93.

[25] See M. Reder, "A Reconsideration of the Marginal Productivity Theory," *Journal of Political Economy* (October, 1947); also TNEC *Hearings*, Part 19, pp. 10503–35, for statements by prominent executives on this matter.

[26] K. Powlison, "The Profit Motive Compromised," *Harvard Business Review* (March, 1950). Mr. Powlison, at the time this article was written, was Controller and is now Secretary, of the Armstrong Cork Co.

The problem is at best a highly complex one and it is possible here only to point out some of the reefs and shoals in the search for the desired profit standard. Unfortunately, it is not a case of looking for a beacon in the dark, because no one measure will suit all purposes best. Two sets of profit standards can be developed, however, that will be of frequent use to managers in guiding their performance: one of these is a group of over-all measures for the firm as a whole; the other consists of internal standards to be employed by the heads of divisional units in the decentralized firm.

Over-all Standards. The choice of an appropriate profit standard for the company as a whole depends on the use to be made of it by management—whether for excluding potential competitors, acquiring capital to finance expansion, maintaining control against creditors, restraining union demands, or other considerations. The following over-all standards may be proposed as relevant criteria.

1. *Comparative Earnings Standard.* The rate of return in comparison with other companies is a commonly used standard by business firms. The measure may take several forms such as the ratio of net income to net worth, the ratio of company profits to industry profits, or the ratio of current profits to profits in some average or normal period in the past (i.e., an index number). Measures of these types have been proposed in antitrust investigations in an attempt to establish that excessively high profits when compared with other firms are a possible indication of monopoly power; labor unions have employed similar measures as an indication of management's ability to pay higher wages. In any event, the use of these measures requires recognition that, especially for industry leaders, the laws of growth will make themselves felt eventually and even the very large company must inevitably accept a declining rate of secular development. Even the industry, as it matures, will be subject to these forces from which there is no escape.

2. *Capital-Attracting Standard.* To finance continued growth may eventually require recourse to the capital markets, and the ability to acquire needed funds can be used as a reflection of the company's status in the investment community where a continued evaluation of alternative investment outlets is always being made. Since capital is obtainable by any established business at *some* price, the mere ability to acquire capital is not in itself a measure of success. If a company is able to acquire capital, it may be necessary to evaluate such financing in terms of dilution of stockholders' equity. For example, the principle underlying the standard might be that the corporation's net profits should, on the average, support a level of market prices for its equity securities such that the issuance of new shares does not reduce the proportionate share of present stockholders in the company's assets, when valued at current prices. In any event, successful financing should result in little or no dilution of the stockholders' equity in terms of the measure employed. Typically, such measures as book value per share, earnings per share, and market value per share are favored. These measures, however, tend to be rather unwieldy because no consistent relationship exists among them, and because much will depend on the type of financing undertaken, especially as between equity and debt.

Where debt is employed, the problem is relatively simple since we may then merely compare the cost of such financing with what seems to be the prevailing rates paid by other established firms in the industry or even in other industries. Where common stock is employed as the financing medium, the acquisition of new funds tends to be tied to some discount (large or small) from current market value. If, then, the stock is selling several times book value (e.g., International Business Machines), it follows that common stock financing will increase the shareholders' book equity. Conversely, where the market price is as much as one-half the book value (e.g., many textile companies) common stock financing will dilute book value per share. This problem is not solved by simply re-evaluating book values in terms of current replacement costs (a process which would be desirable for other reasons) unless we were arbitrarily, and improperly, to raise or lower book values so as to make them coincide with the market prices of the stocks.

Of the measures that might be employed, perhaps the most suitable would be earnings yield. This would involve using an average of, say, the last five years' earnings as a per cent of the price per share received from the sale of new stock. Since the market tends to evaluate the future potential as well as the past performance of the company, this ratio would reflect the investment community's evaluation of the company's position, and the cost it must incur in acquiring equity funds as compared with other capital seekers in the market.

3. *Stockholder Purchasing Power Standard.* Since a business corporation is organized for making profits which will be the source of an income stream for the stockholders, a suitable measure, and one rather

easy to employ, is the *relative* position of the stockholders' purchasing power in the industrial economy. Thus, a company's success may be measured in terms of the growth of dividends per share. If such growth has been sufficient not only to offset purchasing power erosion resulting from inflation but to maintain the *relative* income position of the stockholder as well, e.g., dividend growth at least equal to the growth in industrial wages, then the company has performed quite satisfactorily for the stockholder.

This measure is suitable only for long-run applicability as a profit standard. Applied in the short run it can too readily break down because: (1) many companies maintain the same dividend for a period of time until they feel they can reasonably expect to maintain a higher dividend; (2) companies undergoing rapid growth do not pay out much of their earnings until expansion projects have begun to taper off; and (3) profits are typically unstable and companies particularly subject to earnings instability are likely to permit dividends to fluctuate rather than follow a very difficult stable dividend policy.

4. *Market Appraisal Standard.* Another standard that may be used is to allow a rate of profit for the firm that will preserve the historical relationship between the market price of its equity shares and a broad average of common stock prices. The standard may be a time series expressing the ratio of the price of the stock to the value of the stock index. The latter may be the Dow-Jones Industrial Average or any one of several indexes that are available, or the analyst can devise his own index which for particular purposes might be deemed superior. The approach may also be applied to earnings or to dividends, as well as to market prices, and it might even be the wish of the analyst to compose an over-all index which utilized all three elements in some weighted or unweighted combination. This standard could also be of use in guiding management's plans for long-term growth by providing a favorable setting for capital attraction.

Since market prices, in the long run, tend to reflect the pattern of earnings and dividend growth, the advantage of economy of labor tends to favor series based on price movements rather than a combination of price, earnings, and dividends. On the other hand, since a small, less known company is not likely to find its operation properly reflected in the market price of its stock (or the stock might be closely held so that a realistic market appraisal does not exist) an earnings index would be preferable.

Standards for Internal Use. Determining and employing profit measures for purposes of internal control is an even more difficult and controversial matter than for profit measures for the company as a whole. Economic literature has traditionally concerned itself with the theory of the firm as a profit-making whole, and relatively little energy has been directed, until rather recently, at profit-making subcenters within the firm.

The need for appropriate profit measures (for evaluating performance of subordinate executives and guiding decisions of the subcenter managements) exists in any real sense only in a firm which is decentralized and where management responsibility has been delegated to the heads of the divisional units. In a monolithic organization where all important decisions are made by the firm's top management, the divisional manager is preempted from exercising any discretion over most factors which will shape the profits of his unit, so that the problem of evaluating his performance becomes merely a problem of determining how quickly and how effectively he is carrying out the orders which flow down to him from his superiors at the home office.

The truly decentralized firm is organized as a combination of semiautonomous units and, largely as a result of the fabulous success achieved by General Motors with this type of organization, has won increasing favor among many of our larger manufacturing companies. It has been adopted, among others, by such well-known industrial giants as General Electric, Ford, Chrysler, and Westinghouse Electric. The divisional manager is given authority to plan his selling campaigns, establish selling prices, determine his material and personnel requirements, select his sources of supply either from within or outside of the company as he sees fit, and determine his marketing and distribution channels. Responsibility, in other words, tends to be complete with respect to all short-run decision making. Matters of long-run policy, particularly with respect to capital expenditures, remains the responsibility of the top executive group.

From the above it follows that a profit measure which will function properly should be so designed as to exclude all factors over which the divisional manager has no control. This means that it must be independent not only of the decisions handed down from the top but, as well, from the superior or inferior performance of other divisions with which the one in question "does business."

We can begin to appreciate now why the problem of internal profit measurement is such a difficult one. Many facilities and services may be used jointly by two or more divisions of the company, e.g., general administrative services, research, maintenance personnel, and equipment. Furthermore, one division is likely to use more of these common facilities and services than another, and the amount of such use is not necessarily related to the volume of a division's business, thereby complicating the problem of allocating such costs among the operating units. It is also very likely that one division will have a "business relationship" with another involving a transfer of semiprocessed goods, by-products, and/or services. Where established market prices exist for such product and service transfers the problem is relatively simple; but where no market prices exist, a system of (arbitrary) transfer prices must be established. Since the purchasing division has no control over the efficiency with which such products and services have been produced (and for which no established market prices exist to permit objective testing of their supply prices), it should not be placed in a position of having its own performance hindered or bettered by the supplying division's performance.

The foregoing discussion leads to the conclusion that divisional net profit is a highly unsatisfactory measure to be employed for internal day-to-day decision making and executive evaluation. To this figure should be added back two major cost groups: (1) nondivisional expenses which have been charged to the division as part of its burden for supporting the company overhead; and (2) nonvariable or overhead costs of the division itself, incurred either by decisions made by a predecessor divisional manager or by the top executive group with which rests responsibility for long-term capital commitments made by the division. We then come up with a figure which may be called *controllable divisional profit* and which is essentially the earnings available after deducting from division revenues all variable divisional costs such as materials and administrative and selling expenses, as well as any overhead costs directly subject to the division manager's control.

BIBLIOGRAPHICAL NOTE

A survey and classification of profit theories appears in R. A. Gordon, "Enterprise, Profits, and the Modern Corporation," in American Economic Association, *Readings in the Theory of Income Distribution*, chap. 29 (edited by B. F. Haley and W. Fellner). A generalized uncertainty theory of profit is developed in an article with that title by J. F. Weston, *American Economic Review* (March, 1950), with subsequent critical discussions by several economists in the same journal, March, 1951. Also worth consulting is a recent work by S. Weintraub, *An Approach to the Theory of Income Distribution*, chap. 10. On profit standards, a scholarly work, though with applicability generally limited to companies contemplating war renegotiation contracts, is J. F. Weston and N. H. Jacoby, "Profit Standards," *Quarterly Journal of Economics* (May, 1952). As for break-even charts and related control techniques, most works on cost accounting and on budgeting are quite suitable. See, for example: C. T. Devine, *Cost Accounting and Analysis;* F. V. Gardner, *Profit Management and Control;* T. Lang (ed.), *Cost Accountant's Handbook;* and G. A. Welsch, *Budgeting: Profit Planning and Control.* A penetrating evaluation of break-even analysis from an economic standpoint is Dean's "Cost Structures of Enterprises and Break-Even Techniques," *American Economic Review* (May, 1948), the essence of which appears also in his *Managerial Economics,* pp. 326–41. And on the use of profits for internal control, an approach different from these is G. Shillinglaw, "Guides to Internal Profit Measurement," *Harvard Business Review* (March–April, 1957), where the problem is that of profit control under divisionalization. Finally, on the subject of the separation of ownership and control and its effect on the role of profits, the path-breaking work is A. A. Berle and G. C. Means, *The Modern Corporation and Private Property*, while a more modernized treatment is R. A. Gorden, *Business Leadership in the Large Corporation*, which provides fascinating reading for anyone interested in modern business.

More comprehensive works that are suitable for both supplementary and complementary reading to all or parts of this chapter include: Colberg, Bradford and Alt, *Business Economics*, rev. ed., chap. 1; Dean, *Managerial Economics*, chap. 1; and J. Howard, *Marketing Management*, chap. II. For greater emphasis on policy matters, K. Powlison, "The Profit Motive Compromised," *Harvard Business Review* (March, 1950), may be consulted. Those interested in strengthening their knowledge of the accounting aspects of profit measurement will find a comprehensive treatment in R. D. Kennedy and McMullen, *Financial Statements*, rev. ed., Part III, and in H. G. Guthmann, *Analysis of Financial Statements*, 4th ed., Part I.

QUESTIONS

1. Contrast the three major groups of profit theories as to:
   (*a*) their explanation of the source or derivation of profits;
   (*b*) the distribution of this income share to a factor or factors of production;
   (*c*) the shortcomings of the theory in explaining income distribution.
2. Which theory best describes profits (or losses) resulting from:
   (*a*) du Pont's discovery of nylon;
   (*b*) Polaroid's patent on the Land camera;
   (*c*) regulated public utility profits;
   (*d*) automobile profits after World War II;
   (*e*) railroad operating losses;
   (*f*) a major oil strike opening up a new field;
   (*g*) drug profits from Salk vaccine.
3. As the owner of a fully paid-up home would you—aside from the obvious costs of property taxes, heat, utilities, repairs, etc.—be living in that home "rent free" or would you be incurring various "hidden costs"? What might these "hidden costs" be?
4. Define economic or opportunity cost. Do it in your own words so as to assure yourself (and your instructor) of the full meaning and implications of the concept.
5. Look up and compute the rate of return earned after taxes on the stockholders equity, in 1957, in the following industry leaders: (*a*) U.S. Steel; (*b*) General Motors; (*c*) Standard Oil of New Jersey; (*d*) Aluminum Company of America; (*e*) National Dairy; (*f*) E. I. du Pont; (*g*) Merck.
6 "A baseball game isn't won until the last man is out; an investment isn't profitable until the ownership is terminated." Discuss the significance of this statement in terms of this chapter. What problems present themselves as a result of these implications, and how are they usually met and resolved?
7. Write a short essay on "expectations" as they relate to profit seeking and planning.
8. If a new metal were discovered which had the remarkable properties of never wearing out, rusting, or being otherwise subject to damage or wear, would machines made of this metal be subject to depreciation (accountingwise)?

9. Describe carefully three currently acceptable methods of depreciation accounting, and explain the profit-reporting consequences of each.
10. What kind of enterprises would most greatly benefit by adopting an accelerated depreciation accounting method? Can you indicate any enterprises that would not benefit from such a method? Explain.
11. What is meant by "normalizing" earnings? Illustrate with an example.
12. If a firm were to operate at the "break-even" point, as defined here, would it remain long in business? Why or why not?
13. What is meant by contribution profit? Can it be positive while total profit is negative? Can it be negative when total profit is positive?
14. State the assumptions underlying a typical (linear) break-even chart.
15. Criticize fully the break-even analysis.
16. Distinguish clearly between incremental profit and marginal profit.
17. List five limitational profit factors, and show that one might argue, in each case, that they are only apparently limitational.
18. Discuss three over-all profit standards. Can you suggest two or three others not mentioned in the chapter? Which standard seems best to you?

Chapter 5

DEMAND ANALYSIS—SALES FORECASTING

In the previous chapter we examined the nature of profit and its role in management decision making. As a further step in the process of adjusting to uncertainty, we turn now to the area of demand analysis. Our objective is to discover and measure the forces that affect the sales of a company, and to establish relationships between sales and these controlling forces so that forecasts can be made and thus forward planning facilitated.

The procedure followed in this chapter will be to outline the analytical framework of demand measurement, and to illustrate it by showing the results of various case studies applied to demand measurement and forecasting for a number of different products. In this way the value of demand analysis as a basis for (1) sales and profit budgeting, (2) controlling and manipulating demand, and (3) adjusting production and inventory to future sales expectations, will become more readily apparent.

ANALYTICAL FRAMEWORK FOR DEMAND MEASUREMENT

The theory and measurement of demand, which is the essence of demand analysis, is subject to a number of difficulties both in methodology and interpretation. In any econometric investigation, the nature of these difficulties should be understood if proper use is to be made of the analysis. For the most part the problem consists of bridging the gap between the concept of demand as it exists in economic theory and the measurement of demand by statistical methods. The former, we shall see, provides a guide for judgment, while the latter attempts to yield quantitative estimates within the limits of actual experience.

Demand Concepts: Simple and Multiple Relations

To a professional economist, the term "demand," with reference to market demand, has a specific meaning: it is a dependent or functional relationship revealing the quantity that will be purchased of a particular commodity at various prices, at a given time and place. In elementary economics this relationship is portrayed both arithmetically in the form of a demand schedule, which is a table showing prices and corresponding quantities, and graphically in the form of a chart, which is a pictorial representation of a demand schedule and reveals what is commonly called a demand curve. It should be evident by now, however, that since "demand" is a functional concept, it may be expressed not only arithmetically in the form of a table or graphically in the form of a curve, but also algebraically in the form of an equation such as $D = f(P)$ or $Y = f(X)$; which means that "demand is a function of price" and that a *law of behavior* exists between the two variables. As the equations stand, they state only that a general functional or dependent relationship exists, but they do not tell the exact nature of this relationship. The exact form of demand relationships is a problem of demand measurement and will be treated later in this chapter. Our purpose at the present time is to prepare the groundwork for demand measurement by developing a concept of demand for the solution of practical problems.

Realistically, businessmen know that the demand for most products is affected by many factors other than or in addition to, price. These other factors may include such diverse elements as income levels, the availability of substitute products, advertising and sales promotion, population, availability of credit, season of the year, weather, one's social status, geographic location, and so on. Accordingly, a demand function expressed conceptually in the form of a simple relation as above may be inadequate to explain the variations in demand, and a multiple relation may therefore be necessary. The latter would be expressed as $Y = f(X_1, X_2, X_3 \ldots X_n)$, where each of the X's denotes a specified independent variable, the dots indicate that certain independent variables which may affect demand (Y) have not been specified, and the symbol X_n represents the last specified independent variable. Thus, if we allow price, income, advertising expenditure, and the price of substitutes to stand for the independent variables, and the quantity of coffee purchased as the dependent variable (Y), the above equation would be read, "The demand for coffee (Y) is a function of, or dependent upon the price (X_1), income levels (X_2), advertising expenditure (X_3), certain unspecified factors (shown by the dots), and prices of substitutes (X_n)." The number of demand determinants for most commodities is quite large, and no single set of determinants for any one product is necessarily applicable to another product either in the same combination or to the same extent. Yet, if for any given commodity a few of the most important demand determinants could be isolated, and their joint effects on the total demand for the product could be established, a more comprehensive demand or sales function would be available. A demand function of this type would provide a more general statement as to the nature of the multiple relationship between the dependent variable, sales, and two or more independent variables, such as those stated above. This is precisely what econometricians try to do by the use of correlation analysis, and, as will be seen later in this chapter, by encompassing more than one demand determinant, they develop more comprehensive predicting equations that serve to improve their ability to forecast.

Statistical Considerations

Due to the meaning of the term "demand," the measurement of it can be divided into two sorts of problems. The first is the nature of the price-quantity relation, i.e., the demand schedule or curve, on the assumption that other demand-determining factors remain constant. This type of measurement can be used, for example, as a means for determining elasticity. The second aspect of the problem is to measure changes in the intensity of demand. This type of measurement can be used in determining the nature of shifts in the demand curve. Thus, where management is contemplating a change in price and its subsequent effect on the quantity demanded, it is the first concept of demand in the schedule or curve sense that must be measured; alternatively, if price remains the same and there are changes in other demand-determining factors such as income, advertising, etc., it is the shifts in the demand curve as a whole that are of immediate concern. Realistically, however, in the actual work of demand measurement, these two problem areas are not regarded as mutually exclusive. Analysts are usually concerned both with the nature of the demand curve and with its shifts, for rarely is it possible to measure one without in the same process measuring the other.

Analysts have developed two different methods for making quantitative estimates of demand: one of these involves the use of time-series data; the other is of a cross-sectional nature. Some comments as to both approaches are of interest.

1. Time-series data are sometimes used in which the historical changes in prices, incomes, population, and other variables affecting demand are observed and their interrelationships with demand are measured. Since a demand relation with only certain independent variables is wanted, it may be necessary to eliminate the influence of other independent variables that have a significant effect on demand. Thus, in a demand-price study where the influence of price is the only independent factor under consideration, it is often necessary to make two types of adjustments in the data.

a) *Population Adjustment.* In order to eliminate the effect of population variations on the sale of the product, incomes and demand quantities are reduced to a per capita basis. This adjustment is usually made, however, when the data cover a number of years, since population figures do not usually show sharp fluctuations from year to year. The result of the adjustment is to enable the changes in demand to be attributed to factors other than population. Where the product being analyzed is a family-type good, such as an automobile, washing machine, etc., a better demand estimate is often obtained by reducing the relevant data to a per household rather than per capita basis. In any event, it should be realized that such reductions do not, of course, adjust for changes in the age distribution, racial composition, or other elements in the population that may affect demand over the long run.

b) *Deflation Adjustment.* A similar reduction, usually called "deflation," is to adjust for changes in the purchasing power of money by dividing the price series in current dollars by an average price index of all goods. An example of the latter and one that is commonly used in consumer demand studies is the Consumer Price Index, since it reflects the average prices paid by consumers for most goods and services. Although this procedure yields fairly satisfactory results, it should be kept in mind that deflation methods of this sort do not give precise measures of price changes mainly because no perfect index has yet been constructed and because the time period covered may be too long.[1]

2. Cross-sectional analysis attempts to discover how consumption by individuals or families varies with prices, incomes, geographic differences, and the like, at the present time rather than over a period of time. This is similar in many ways to a controlled experiment, in that variations in the data are current and not historical. For example, in establishing a sales-income relationship for the purpose of measuring the income elasticity of demand (discussed later in this chapter), the time-series method would employ past variations in the data as a basis for measurement. The cross-sectional method, on the other hand, would compare the different levels of sales at the present time among different income groups, and the elasticity measure would be based on these differences. But as in the time-series method, adjustments in the data may also be needed in order to eliminate the effects of other factors (in this case all factors other than income) that may affect significantly the demand for the product. In any event, the choice of either approach depends upon time and expense considerations, and the data already available. For these reasons, the time-series method is more commonly employed—the data already available from company records—with minor use made of cross-sectional information when it seems appropriate.

Simple Versus Multiple Relations

A consideration of practical importance concerns the question of whether the predicting equation to be derived should be based on a simple correlation between a dependent variable, in this case demand, and one independent variable, such as price, or whether it should be a multiple correlation involving two or more independent variables, e.g., price, income, and other demand determinants as controlling factors. Simple relations have the advantage of being easier to compute; from the management standpoint they are also easier to comprehend and easier to manipulate, since only one controlling variable is involved. But a demand analysis involving a simple relation may be of limited application, since the attempt is to manipulate demand by varying this single controlling factor. The forecasting reliability of the function may also be questioned: if disposable income is the independent variable, it will be correlated with broad product groups and there may actually be little or no causal relation between it and the specific product under analysis. When price, on the other hand, is the independent variable, it will likely be a controlling factor only in the short run while other demand-determining elements are still relatively constant. Multiple correlation, however, permits the introduction of several independent variables as controlling factors, and for many kinds of practical problems only a few such variables are necessary to explain the great majority of the variations in the dependent variable. Multiple correlation, therefore, provides for a more general demand function as compared to the results achieved by using only simple relations. But it is also more expensive and, in terms of the extra time and expense, may not always be worth the gain in precision. As pointed out in the discussion of correlation in Chapter 3, analysts are not usually concerned with accounting for all of the variations in the dependent variable, but rather the great majority (say 90 per cent) of those variations. Therefore, simple correlation may often be quite adequate. The applications of both simple and multiple methods, however, will be presented in the following sections.

---

[1] In addition to the population and deflation adjustments in time-series analysis, other adjustments are also sometimes made, such as removal of trend, seasonal, and cyclical influences. Two procedures employed for such purposes, discussed in Chapter 3, are: (1) calculate "normal" values and then express the original data as percentages of, or as deviations from, these normal values; or (2) use first differences or link relatives of the original data. It follows, of course, that once any fluctuations are removed from the data they are no longer observable in the original series.

DEMAND DETERMINANTS—ELASTICITY

The demand for a Webcor hi-fi set is determined by a great many factors, among the most important of which are (1) its price, (2) buyers' incomes, (3) the price of available substitutes or competing products, (4) Webcor's advertising, (5) the availability of credit, and (6) perhaps other factors such as geographic location of buyers, their economic expectations, and so forth. Obviously, it would be impossible in most demand studies to include all of the factors that exert an influence on sales. For a great many products, however, the first four of the above factors—price, income, substitutes, and advertising—expressed and measured in various ways, have the greatest influence on sales, and they are the controlling variables commonly used in many if not most demand studies. Accordingly, we shall devote the present section to discussing and illustrating their use in the measurement of demand. The procedure followed for the most part will be to present the graphic results of econometric studies employing simple correlation, reserving multiple correlation studies for the following sections.

Since we shall be dealing with the actual measurement of demand, first in terms of simple relations, the demand function may now be expressed conceptually in a manner somewhat more refined than that stated earlier. That is, we recognize that demand is a function of several variables, but all of these except one are to be held constant in establishing an actual relationship. Accordingly, the general nature of the studies to be discussed in this section are of the form $Y = f(X_1 \mid X_2, X_3 \ldots X_n)$, where the vertical bar signifies that the causal factors to the right are to be regarded as fixed in the demand function under analysis, the factor to the left is to be varied, and the dots as always indicate that certain causal factors have not been specified. For measurement purposes, however, when only one factor is varied, the equation can be written simply $Y = f(X)$, since the factors assumed to be held constant are not included in the actual demand function which is eventually derived.

One further comment, this time as to the practical significance of the empirical studies to be discussed, is in order. Throughout this and subsequent chapters dealing with the actual measurement of economic relationships, emphasis will be placed on the importance of various elasticity measures. Elasticity may be defined in general as the percentage change in a dependent variable resulting from a 1 per cent change in the independent variable. Such measures may be derived from curves or from estimating equations, as will be seen shortly when specific measures such as price elasticity, income elasticity, and the like are discussed. From the general definition of elasticity, it should be evident that such measures are a guide to improved prediction, and hence a means of reducing the uncertainty inherent in forward planning by management.

Price

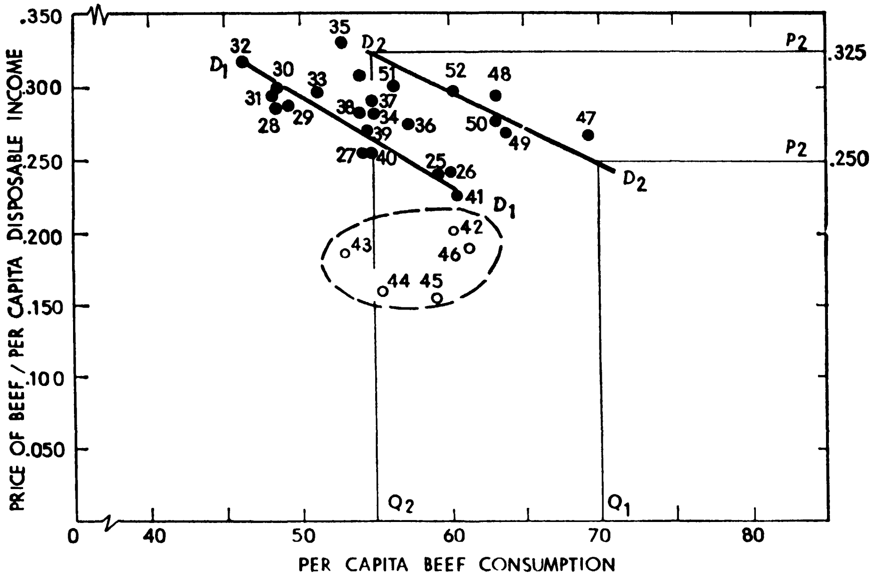
The relation of price to sales has been a major interest of economists for a long time. A better knowledge of such relations is also of concern to management, however, as a basis for pricing, demand manipulation, and profit control. Despite these advantages, empirical studies of pricing have been relatively rare in manufacturing, as compared to agriculture. In the latter sector of the economy there are wide variations in prices and the effects on consumption are often discernible in the short run; in manufacturing, on the other hand, prices remain stable for long periods and the effects of their changes are usually combined with general business conditions, thereby complicating still further the problem of separation and measurement. Hence, to a large extent, controlled experiments (experimental designs) offer a promising approach to the study of short-run sales-price relations, particularly where manufactured goods are concerned. In the discussion below, the results of three demand studies in the agricultural, manufacturing, and service sectors of the economy are outlined as an illustration of the sort of price-sales relations that are relevant from the management standpoint.

Demand for Beef. Figure 5-1 illustrates some results of a demand study for beef. The data cover the period from 1925 to 1952. From the appearance of the dots it seems that there are actually three separate patterns of price-quantity relationships, one for the prewar period, one for the war period, and one for the postwar period. The prewar scatter is represented by a least-squares regression line, D_1D_1, which is the demand curve for beef for the period 1925 to 1941. During the war years, price ceilings and rationing depressed beef consumption, so the data for these years have been excluded from the analysis and encircled in a dashed line. With the removal of wartime controls, conditions returned to normal and a new demand curve, D_2D_2, is discernible for the period 1947 to 1952. The analysis thus reveals that there has been an increase in demand—a shift of the demand curve to the right from D_1D_1 to D_2D_2, with beef consumption running about 20 per cent higher after the war than before.

FIGURE 5-1

DEMAND FOR BEEF

Price of Beef Divided by Per Capita Income, Plotted against Per Capita Beef Consumption, Annually, 1925-52



Source. Adapted from G. S. Shepherd, *Marketing Farm Products*, 3d ed., p. 310. Reprinted by permission of the Iowa State College Press.

The chart also reveals the elasticity of the demand for beef—that is, the responsiveness of quantity demanded to a change in price. The demand elasticity (E_D) may be readily computed from the formula

$$E_D = \frac{\dfrac{Q_2 - Q_1}{Q_2 + Q_1}}{\dfrac{P_2 - P_1}{P_2 + P_1}} \substack{>\\=\\<} 1$$

where Q_1 and Q_2 represent the quantity demanded before and after the price change, respectively, and P_1 and P_2 are the prices that correspond with these quantity figures. Thus, in measuring the demand elasticity on D_2D_2, the purpose is to determine the elasticity for the curve as a whole. Therefore, the vertical and horizontal lines in Figure 5–1 should be drawn to or near the ends of the demand curve, as shown. If the elasticity over a smaller price range is desired, the P_1P_2 band (and hence the Q_1Q_2 band) can be narrowed accordingly. From the above formula, the elasticity is about 0.9, which means that demand is slightly inelastic.

$$E_D = \frac{\dfrac{55 - 70}{55 + 70}}{\dfrac{.325 - .250}{.325 + .250}} = \frac{-.12}{.13} = 0.92$$

The minus sign is understood since price and quantity demanded are inversely related, and hence the sign may be omitted from the final answer. Also, it should be noted that the choice as to which end of the curve to designate as P_1Q_1 and P_2Q_2 is immaterial, since the same answer will be obtained either way. Interpreting the result, the elasticity coefficient means that a 1 per cent increase (or decrease) in price may be expected to bring about 0.9 per cent decrease (or increase) in quantity demanded, other things remaining the same. Similarly, a 10 per cent change in price will be associated with approximately a 9 per cent opposite change in quantity demanded. Elasticity measures are thus a useful tool for better prediction and planning by management. The derivation of such measures is often a chief purpose of empirical studies not only in demand analysis, but also in other areas of managerial economics as will be seen in later chapters.

Several graphic methods can be employed to measure elasticity, one of the more common being to plot the original data on double logarithmic paper, i.e., paper on which both axes are scaled logarithmically (or, what is the same thing, plot the logarithms of the data on ordinary arithmetic paper). Then, regardless of the units in which the data are quoted, such as dollars, pounds, etc., the elasticity can be determined directly from the chart by measuring with a ruler the change in quantity, ΔQ, and the corresponding change in price, ΔP, for the range of the curve desired. The elasticity would then be equal to the ratio $\Delta Q / \Delta P$. The reason for this is that by using logarithms, the scales of the chart are automatically converted such that equal distances on both axes represent equal percentage changes, and elasticity is, after all, a measure of relative change between the dependent and independent variables. It follows, therefore, that if quantity is plotted *vertically* and price *horizontally* on log paper, the elasticity of the curve would equal its slope, for the latter is always the number of units that the curve rises vertically, ΔY, per unit of horizontal run, ΔX, or $\Delta Y/\Delta X$. In other words, elasticity equals slope, or $\Delta Q/\Delta P =$

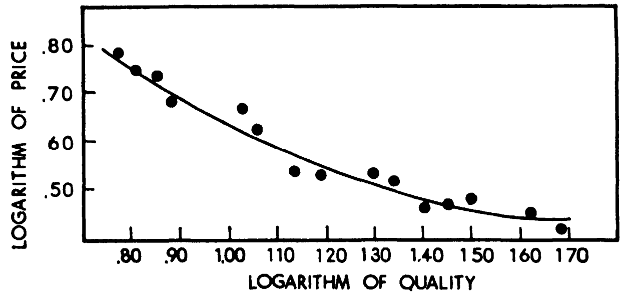
$\Delta Y/\Delta X$. But economists traditionally plot price vertically and quantity horizontally, so that, technically speaking, elasticity is not the same as the slope but rather the reciprocal of the slope:

$$\Delta Q/\Delta P = \Delta X/\Delta Y$$
$$\text{(elasticity)} = \text{(reciprocal of slope)}$$

Therefore, the greater (steeper) the slope the less the elasticity. This is a convenient concept to keep in mind when judging visually the elasticity of any curve plotted on (double) logarithmic paper where the independent variable is scaled vertically and the dependent horizontally. When the plotting is reversed, however, as is typical in the graphing of all other functional relationships in economics (discussed below), the dependent variable is scaled vertically and the independent horizontally.

FIGURE 5–2

DEMAND FOR A MILLINERY PRODUCT
A Department Store Study



In such cases the elasticity can be judged directly from the slope of the line rather than from its reciprocal. The greater (steeper) the slope in such instances, the greater the elasticity. This notion will become clearer later in this section where income, substitutes, and advertising are discussed as controlling factors in demand measurement, and are plotted horizontally as independent variables.

Demand for Millinery. In contrast to the previous study in which the data were derived from a time series covering many years, Figure 5–2 shows the demand curve for a millinery product derived for a leading Detroit department store. The data were obtained from controlled experiments covering a period of 15 shopping days (Tuesdays to Thursdays) over five weeks. The short time period allowed for wide price manipulations while other influential factors, such as income, fashion, and competitors' reactions, could be assumed to remain constant. This made it possible to confine the analysis to one of simple correlation between demand and price. Since the logarithms of the data have been plotted, it is immediately apparent from the slope of the curve that the elasticity is greater than unity, i.e., that the curve is relatively elastic.

Measured graphically, the coefficient of elasticity is about −3, meaning that on the average, a 1 per cent increase in price is associated with about a 3 per cent decrease in unit sales. Though this coefficient may seem high, it is not as large as other measures that were derived when the experiments were applied to other, more staple products in the store. For some products the coefficients ranged between −5 and −10, probably because they were in greater competition with similar products carried by other stores in the area, and price, therefore, was a more significant demand determinant.

Studies of this type are illustrative of the sort of price-sales analyses that can be made in the area of demand measurement by the method of controlled experimentation. Although, realistically, prices of most manufactured goods do not ordinarily change in the short run but instead remain stable sometimes for months or even years, this approach as a method of analysis shows promise of gaining increasing recognition in the future. As yet, however, the study of demand (price-sales) relations of the type illustrated above is still a relatively untouched area by analysts outside of the field of agricultural economics. Until now, much of the pioneering work in demand measurement has been done by researchers interested in food marketing and related problems, and a good deal of what is known about the techniques of demand measurement stems from studies done in the food field.

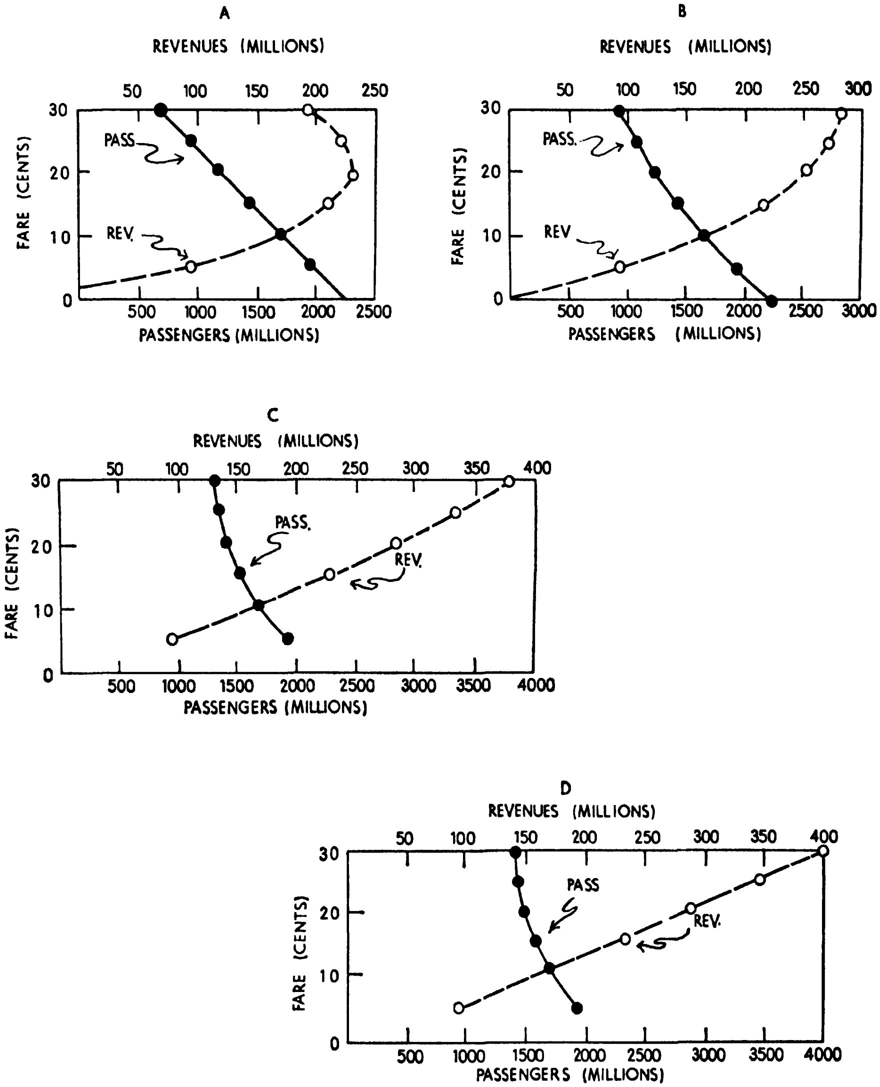
Demand for Subway Travel. William Vickrey of Columbia University made a study in 1952 of the fare structure and its effects on passenger travel for the public transit system of New York City. He made projections of demand and revenues based on various fare levels, the results of which are shown in Table 5–1 for subway service. Figure 5–3 presents the same information graphically as shown in the table. In Chart A of Figure 5–3, the assumption is made that equal *absolute* changes in fare produce equal *absolute* changes in traffic, so that the passenger demand curve is linear. Total revenue as shown by the dashed line is maximized at a fare of 20 cents. (It may be helpful when looking at the revenue curve to rotate the page by a half-turn counterclockwise so that revenues appear on the vertical axis and fare on the horizontal. The maximum revenue point, about $232 million, is then seen to be directly above the fare of 20 cents when passenger traffic is about 1,160 million.) At fares higher or lower than this, total revenue is below the maximum.[2] Chart B is based on the exponential assumption that equal *absolute* changes

---

[2] Where the total revenue is a maximum, i.e., neither increasing nor decreasing, the corresponding demand elasticity at the point is unity. For at this point, a small change in price must result in an exactly proportional change in quantity in order for revenue to remain constant. In practical work, the elasticity is usually measured over a range or arc of the curve and the resulting coefficient is an average estimate. This is the nature of the distinction between *point* and *arc* elasticity commonly made in economic theory.

FIGURE 5-3

Demand for Subway Service: Alternative Projections
Passengers and Revenues in Millions per Year



Source: Table 5-1.

in fare produce equal *percentage* changes in traffic. Chart C expresses the logarithmic assumption that equal *percentage* changes in fare result in equal *absolute* changes in traffic. And finally, Chart D shows that on the assumption of constant elasticity, equal *percentage* changes in fare yield equal *percentage* changes in traffic. Economically, Charts C and D are absurd at low fare levels, since they imply an infinite amount of traffic at a zero fare; A and C imply a fare at which all traffic would be suppressed, while B and D imply that there would be some traffic despite a high fare. Vickrey concludes that on the whole, B perhaps represents the most reasonable pattern for the range of fares considered, since it implies that traffic at a fare of 25 cents would be about half that with free service and a little over two-thirds what it was at 10 cents.

Studies of this type can become highly elaborate as more complex circumstances are taken into account. In arriving at a decision as to the optimum fare, many other factors had to be taken into consideration, such as the significance of short- and long-haul riders, the influence of substitutes (buses, car pools, taxis, or even walking), and the time of day (e.g., special fares for rush hours), to mention only a few. Nevertheless, analyses of this sort as well as those illustrated previously are indicative of the kinds of demand studies management needs as a guide to better prediction and planning where price-sales relations are involved. Similar analytical techniques are also applicable to other areas of business economics as will be brought out below and in subsequent chapters.

TABLE 5-1

COMPARISON OF ALTERNATIVE PROJECTIONS OF THE DEMAND CURVE FOR SUBWAY SERVICE

Projected Passenger Traffic and Revenues, Millions per Year, 1950 Level

FARE (Flat)	LINEAR DEMAND CURVE (a)		EXPONENTIAL DEMAND CURVE (b)		LOGARITHMIC DEMAND CURVE (c)		CONSTANT ELASTICITY (d)	
Cents	Passengers	Revenues	Passengers	Revenues	Passengers	Revenues	Passengers	Revenues
0 ...	2,207	$000 00	2,248	$000.00	Inf.	$.	Inf.	$.....
5....	1,945	97.2	1,945	97.2	1,945	97.2	1,945	97.2
10....	1,683	168.3	1,683	168.3	1,683	168.3	1,683	168.3
15 ...	1,421	213.2	1,458	218.7	1,530	229.5	1,547	232.0
20....	1,159	231.8	1,262	252.4	1,421	284.2	1,457	291.5
25 .	897	224.2	1,092	273.0	1,347	336.8	1,390	348.2
30 ...	635	190.5	945	283.5	1,278	383.4	1,340	402.0

Source W. S. Vickrey, *The Revision of the Rapid Transit Fare Structure of the City of New York*, (Tech. Mon. No. 3, Finance Project, Mayor's Committee on Management Survey of the City of New York), p 87.

Income

The income of buyers is a basic demand determinant and, along with price, often accounts for most of the variation in sales for many marketable products. A simple relation of sales to income is of use to businessmen in planning sales, allocating territories, and the like. The relation of total consumption expenditures to total income, called the "consumption function" and studied in elementary economics, has occupied the attention of economists for some years, and some general conclusions have emerged that are useful as a basis for directing management thinking along the appropriate lines. The paragraphs below, therefore, will describe first, the over-all characteristics of the total consumption function; second, the measurement of specific sales-income relations, or what might be called "product consumption functions," and third, the significance of regional differences in income as a basis for establishing sales-income relations. These represent three separate facets of analysis in using income as an independent variable in demand measurement.

Consumption Function. The late J. M. Keynes fashioned the relationship of consumption to income as a tool of analysis, based on his "fundamental psychological law" that consumption changes with income, but more slowly than the latter. A number of economists, notably Duesenberry, Kuznets, Modigliani, Friedman, and Samuelson, have analyzed empirically the validity of this statement and its ramifications for data covering the 1920's and 1930's, and some of their conclusions can be stated briefly as follows.

1. The data analyzed, covering a period of several decades, reveal that the long-run relation of consumption to income is somewhat stable, and that consumption expenditures are regularly about 85 to 90 per cent of income.

2. For short-run, quarter-by-quarter analyses, the results show great instability between consumption and income, and the relationship cannot be predicted by any mathematical formula.

3. During the upswing of a business cycle, consumption expenditures tend to increase in absolute amounts, but decrease as a percentage of income, while in the downswing, consumption decreases in absolute amounts but increases as a percentage of income. In other words, total consumption expenditures are a larger percentage of income in depression periods and a smaller percentage of income in times of prosperity. Savings, of course, since they are the difference between income and consumption, are a larger percentage of income in prosperity and a smaller percentage in depression.

4. In the long run, an individual's spending habits are based on the distribution of income within his community, his place within that distribution pattern, and his desire to emulate the consumption of others (i.e., keep up with the Joneses). Therefore, as long as the pattern of income distribution remains about the same, the proportion of consumption expenditures to total income, called the "average propensity to consume," will remain fairly stable.

5. Finally, it is easier for consumers to raise their standard of living than it is for them to lower it. Therefore, the rate of consumption increase is greater in periods of revival than is the rate of consumption decrease in periods of recession. When the economy is in a downswing, consumers try to maintain their standard of living in the face of falling incomes, and thus consumption expenditures become a larger proportion of income. Cultural lags, of course, are also a dominant factor in the analysis of consumption-income dynamics. When a family experiences a

change in income, it may take a substantial period of time before it adapts to the new income level. Until the full adjustment is made, its consumption habits and patterns may be substantially different from the average for its income group.

The above characteristics of the total consumption function introduce obstacles to the accurate prediction of consumption. Yet, total consumption exhibits a more stable relation to income than do broad product groups, e.g., durables and nondurables. For, given the decision to spend, the choice of how to spend depends on relative prices, consumer stocks, and other factors as well as income. Nevertheless, despite the less predictive feature where product groups are concerned, some useful product consumption functions can be derived for estimating sales-income relations.

Product Consumption Functions. Figure 5–4 illustrates the income-sales relation, or product consumption function, for six product groups, and also the shift in the regression lines that occurred between the prewar and postwar periods. The regressions were derived by correlating dollar expenditures for each item with disposable income over the periods 1929–40 and 1947–54. It should be noted that the correlation equation used is linear in logarithms with disposable income as the independent variable. A measure of income sensitivity can easily be derived, therefore, by simply measuring the slope of the lines with a ruler. For furniture, the prewar coefficient of income sensitivity was 1.6 while the postwar coefficient was 0.6. Interpreting the latter, this means that a 1 per cent change in disposable income (the independent variable) is associated with a 0.6 per cent change in furniture sales (the dependent variable).[3] In a similar manner, the income sensitivity coefficients for the remaining products in Figure 5–4 can be measured graphically and the results interpreted accordingly.

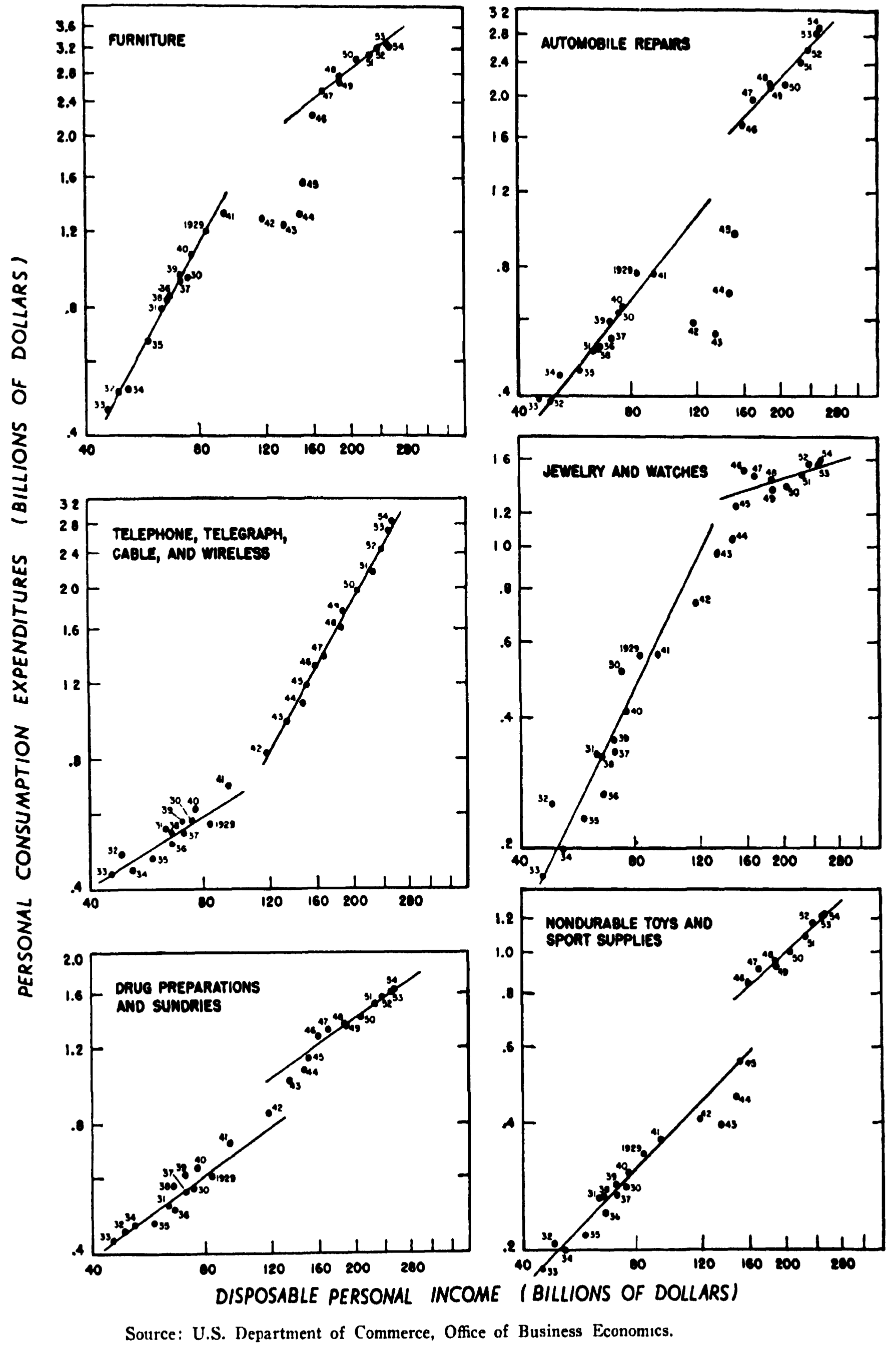
As a benchmark for forecasting, the lines in Figure 5–4 can be extrapolated for projected income levels and adjusted for the trend in the deviations of the dots from the regression line. This assumes, however, that the established relationship is durable enough to hold for the forecast period. Often, the narrower the product line (e.g., couches as compared to furniture), the more difficult it is to fit a meaningful regression line to the data. The method has the advantages, however, of being susceptible to freehand analysis, is relatively inexpensive, and is easily comprehended by management.

---

[3] This measure of income sensitivity is not quite the same thing as the income elasticity of demand. In the latter case quantities purchased are used and the income elasticity is derived from an equation involving major demand factors such as income and price. Income sensitivity, on the other hand, measures the per cent change in dollar expenditures associated with a given per cent change in income, other factors being equal. It thus reflects the influence of income on consumption. However, to the extent that other factors are correlated with income, their effects will also be reflected by the income sensitivity coefficient.

FIGURE 5-4

Examples of Shifts from the Prewar to the Postwar Relationship between Expenditures and Income



Source: U.S. Department of Commerce, Office of Business Economics.

Regional Incomes. Businessmen whose markets are restricted to certain areas of the country are usually more interested in market studies that reveal the purchasing power of specific regions. A commonly used guide for such purposes is an index of buying power published by *Sales Management* magazine for states and counties in the United States, and is a standard reference source for marketing researchers. In 1946, however, the Commerce Department made a study of the extent to which changes in state incomes are associated with changes in national income payments over a period of years. Their investigation covered the years 1929 to 1940, inclusive, and although the same estimates may not be applicable today, a few of the conclusions will be presented in order to illustrate the type of analysis employed. Once the study is understood, the results can readily be brought up to date by utilizing more recent data.

Figure 5–5 shows the remarkably close relationship that exists between state income payments and the nation's total, particularly with respect to the direction of movement, e.g., Chart A for Ohio and the United States as a whole. The relationships may be observed more closely, however, by plotting the data on scatter diagrams as done in Charts B, C, and D for states chosen at random. The figures have been plotted on ratio scales (double logarithmic paper) in order to make the percentage changes in income for the state and for the nation comparable. Formulas expressing the relationships were also derived. The coefficient of log X represents in each case the sensitivity (elasticity) index. Thus, for Ohio, a 1 per cent (or 10 per cent) change in national income payments is associated with approximately a 1.1 per cent (or 11 per cent) change in Ohio income payments. The same techniques were also applied to entire regions of the country, e.g., New England, Middle Atlantic, East North Central, etc., and income sensitivity measures were thus derived.[4]

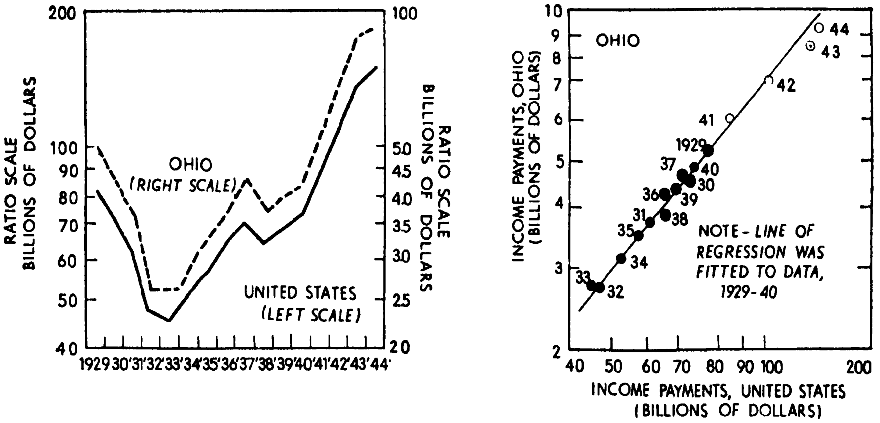
Studies of this type illustrate the sort of use that can be made of readily available income data as a basis for better sales planning and forecasting on a local or regional level. The techniques are relatively simple and the results can be derived either graphically or mathematically. For many products, a functional sales-income relation is relatively easy to determine. More important is the choice as to the type of income aggregate to select, which in turn is determined by the nature and purpose of the analysis and the information available.

---

[4] In a number of states a simple direct relation between state and national income payments was not sufficient to explain all the variations. In twenty states, for instance, a definite downward or upward time trend was observed after the effects of changes in national income payments were eliminated. Accordingly, the additional factor of time was introduced to take care of the trend variations for these states. In the case of New York, for example, the regression equation came out to be $\log Y = 2.502 - 0.0061t + 0.850 \log X$, where Y represents state income payments in millions, X is nation's income payments in billions, and $t =$ year—1935.

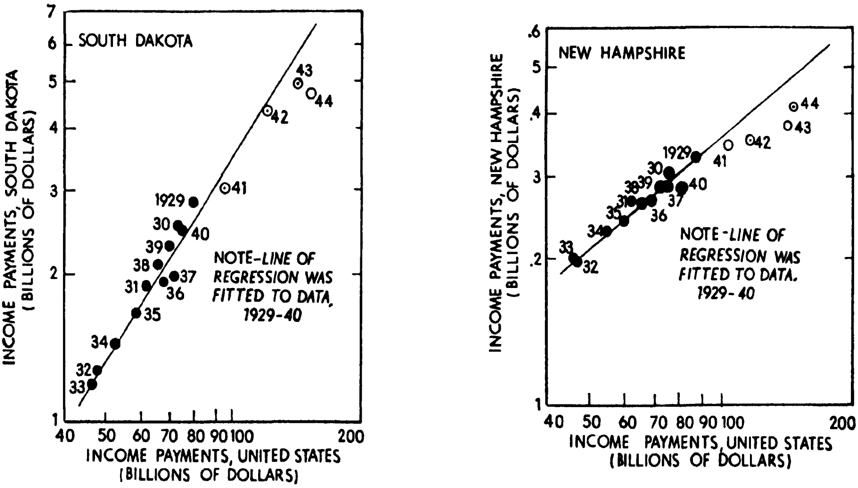
FIGURE 5-5

Relationship between Income Payments for Specified States and the United States



A.

B LOG Y = 1 579 - 1 103 LOG X



C. LOG Y = 1 172 - 0 679 LOG X

D. LOG Y = -0 302 - 1 429 LOG X

Source: *Survey of Current Business* (January, 1946).

Substitutes

Commodities may be related from the standpoint of demand in any one of three ways. They may have a competitive relationship, in which case they are *substitutes* for one another in the consumer's expenditure plan; they may be *independent;* or they may be *complementary.* Commodities are substitutes when an increase in the purchase of one is at the expense of a purchase of the other, as when a family buys a Ford instead of a Chevrolet or Plymouth. Commodities are independent when the purchase of one has no direct influence upon the demand for the other. Examples of such products are numerous, although it can be argued that out of any given income, all products stand in competitive relation with one another and with saving as far as the buyer is concerned. For purposes of demand measurement, however, the relevant criterion is whether the product purchased has a direct (and usually immediate) effect on any other product such that the relationship can be justified economically. Finally, commodities are in complementary demand when an increase in the purchase of one causes a rise in the consumption of the other. Examples here are strawberries and cream, pizza and beer, lamps and tables, automobiles and service stations. There are, of course, degrees of substitutability and complementarity: some products may have a one-to-one relationship, e.g., watches and watch bands; others may vary in widely differing ratios, e.g., shirts and ties. The broader the variation in the ratios the more difficult it may be to determine the competing or complementing effects for prediction purposes.

Several techniques exist for measuring the degree of substitutability among products, but two of the more common ones will be illustrated here.

1. The *cross elasticity of demand* measures the percentage change in the consumption of product Y resulting from a 1 per cent change in the price of product X. In symbols, letting P_x represent the price of X, the formula is:

$$P_x{}^{E_y} = \frac{\dfrac{Y_2 - Y_1}{Y_2 + Y_1}}{\dfrac{P_{x2} - P_{x1}}{P_{x2} + P_{x1}}} \gtreqless 0$$

The data for the calculations below are derived from Figure 5–6 which represents a simple regression of margarine consumption on butter prices for the period 1924–41. The cross elasticity is measured in the same manner as the direct price elasticity illustrated earlier in Figure 5–1 for beef consumption. In this case the coefficient of cross elasticity comes out to 1.1:

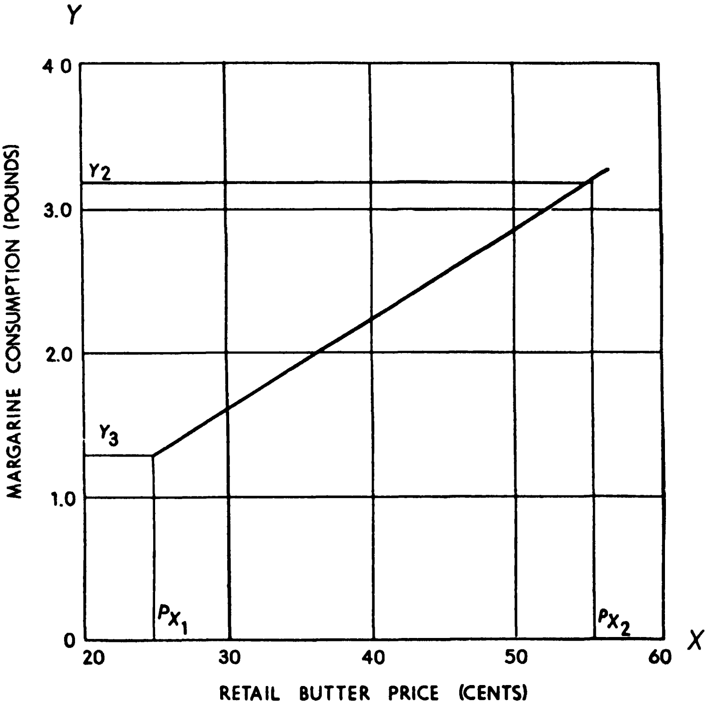
$$P_x{}^{E_y} = \frac{\dfrac{3.2 - 1.3}{3.2 + 1.3}}{\dfrac{55 - 25}{55 + 25}} = \frac{\dfrac{1.9}{4.5}}{\dfrac{30}{80}} = 1.1$$

That is, for the period covered by the data, a change of 1 per cent in the price of butter is associated with an approximately proportional change in the per capita consumption of margarine. Since the curve slopes upward, it means that the products are directly substitutable: higher butter prices occasion more margarine (and less butter) consumption. A horizontal curve would give a zero elasticity coefficient indicating that the two products are independent. A negatively (downward) sloping curve would indicate complementarity: higher prices of X result in a smaller consumption of both X and Y. Of course, the data could have been plotted on double logarithmic paper in which case the elasticity could have been determined graphically by measuring the slope of the line.

2. The *elasticity of substitution* is a second measure of substitutability. It has been defined in production economics as "the proportionate change in the ratio of the amounts of the factors employed divided by the proportionate change in the ratio of their prices to which it is due."[5] "It represents the additional amount of the factor B, from the given combination of factors, necessary to maintain product unchanged when a small unit reduction is made in the use of factor A."[6] The same definition is applicable in principle when measuring demand. In Figure 5–7, the elasticity of substitution is measured between three competing brands of sporting equipment. The data were derived from controlled experiments as part of the department store study mentioned earlier. The

FIGURE 5–6

RELATION OF MARGARINE CONSUMPTION TO RETAIL BUTTER PRICES, UNITED STATES, 1924–41



---

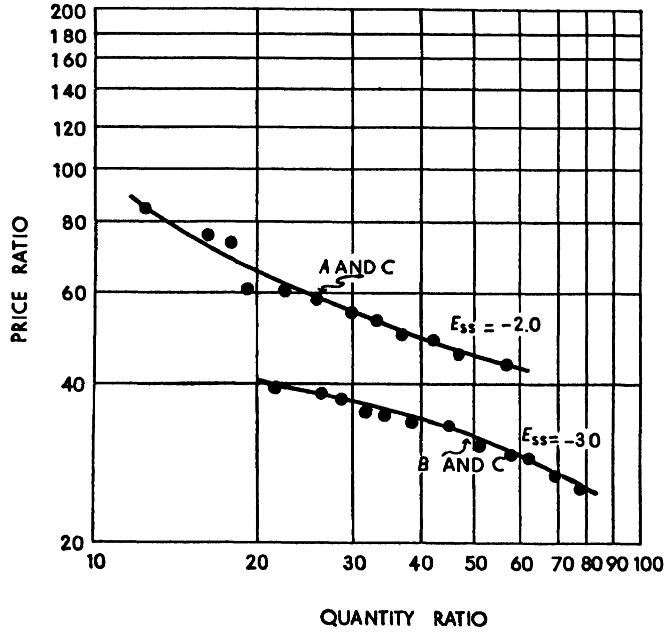
[5] Joan Robinson, *Economics of Imperfect Competition*, p. 256.

[6] R. G. D. Allen, *Mathematical Analysis for Economists*, p. 341.

elasticity of substitution (E_{ss}) between brands *A* and *C* (upper line) appears to be about −2.0, while that between *B* and *C* (lower line) is about −3.0. This means that consumers found it easier to substitute *B* for *C* than *A* for *C*. Note that since the data are on ratio paper, the E_{ss} coefficient can be obtained directly by measuring the average slope of the line with a ruler. However, since the dependent variable, the quantity ratio, is plotted horizontally and the independent variable, the price ratio, is plotted vertically, the E_{ss} coefficient in this case is the reciprocal of the slope, or simply $\Delta Q/\Delta P$.

FIGURE 5-7

ELASTICITIES OF SUBSTITUTION BETWEEN THREE COMPETING BRANDS OF SPORTING GOODS



The idea of substitution and the corresponding elasticity measures are widely used in economic theory. Relatively few statistical studies have been done, however, for the purpose of deriving numerical measures. An understanding of such relationships can be useful to management for price planning and demand manipulation. But as mentioned earlier, the data must usually be obtained from controlled experiments (especially where manufactured goods are concerned) and the process may prove costly if not dangerous, particularly where oligopoly markets are involved and opportunities for open price manipulation are thus limited.

Advertising

For purposes of demand measurement, advertising is viewed as a kind of selling cost or outlay the function of which is to increase sales by shifting the firm's demand curve to the right and upward. Advertising (i.e., selling cost) is thus distinguished from production cost in this way: *selling costs* affect sales, are a cause of sales, and are incurred for the purpose of influencing the buyer in his choice of product and seller; *production costs*, on the other hand, are those arising from the actual production of the goods. Included in selling costs, therefore, are not only the costs of transactions, e.g., the expenses of buyer and seller communications, order taking, etc., but also the expenses of dealings that take place between companies that are faced with rational rather than emotional buying appeals. (Even locomotives are sold by salesmen.) Advertising is thus a device for altering sales volume, by shifting the demand schedule to a higher level.[7]

Theoretical Model. To facilitate analysis, a sales-advertising model can be constructed which portrays the salient features of an operating system and indicates the kind of measures that are needed. Since advertising is viewed as producing sales, the model is a type of advertising "production function" as shown in Figure 5–8, the fuller meaning of which will become more evident in the following chapter where the concept of a production function is examined more closely. We can, however, establish some fundamental ideas at this time.

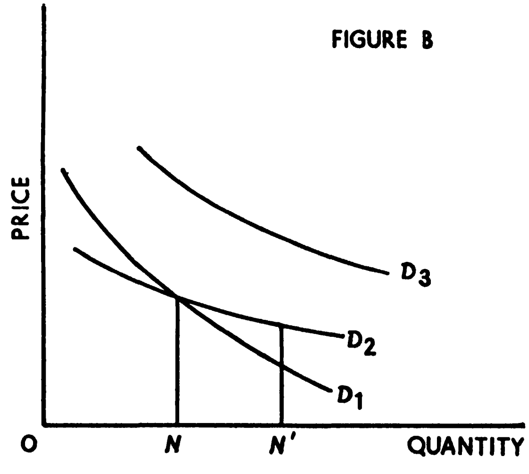
Basically, there are two questions to be answered: (1) How much should a company spend for advertising, or in other words, what should be the size of its advertising budget? (2) Given the size of its advertising budget, how should the company go about allocating its expenditure among competing media? The first question relates to the measurement of demand and is discussed below. The second involves the subject of cost analysis and is treated later in Chapter 7.

In Figure 5–8, advertising expenditures are measured horizontally and the resulting sales, costs, and net profits vertically. The rising diagonal line represents advertising costs, and the curved line labeled gross profits

---

[7] It is often said that a desirable effect of advertising, from the seller's standpoint, is that it makes the demand curve more inelastic, thus allowing a higher price to be charged. If this means (Fig. B) that D_1 is preferable to D_2, it is false. If the most profitable output is beyond *ON*, say at *ON'*, D_2 is preferable to D_1 because it allows for sales at a higher price, even though D_2 is more elastic than D_1. Actually, what the seller wants is a higher curve *level*, such as D_3 relative to either D_1 or D_2.

FIGURE B

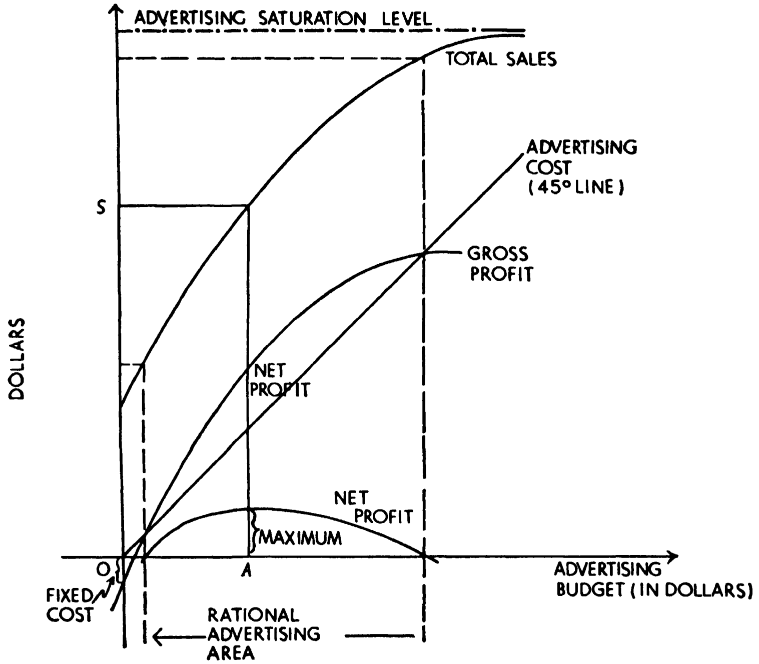


represents the difference between sales and all costs except advertising. Net profit is thus the area between gross profit and advertising costs, and the net profit curve when plotted separately takes the inverted "bathtub" shape as shown in the diagram.

The analysis assumes that price, quality, media, and other factors that may affect sales are held constant, so that different sales levels can be attributed to variations in advertising expenditures. The fact that sales are some positive amount even when advertising is zero indicates that a certain amount of sales are forthcoming even without advertising. With successive doses of expenditures on advertising, the sales curve rises, eventually at a decreasing rate, indicating that diminishing returns to advertising are setting in. That is, each increment in advertising, say in $1,000 expenditure blocks, produces a less than proportional increase in sales, so that the ratio of the change in sales (ΔS) to the change in advertising expenditure (ΔA) decreases. The ratio $\Delta S/\Delta A = 0$ where total sales are a maximum, and it is seen that this occurs beyond the advertising expenditure level at which net profits are maximized. Evidently, the optimum advertising budget, if maximum net profit is the criterion, is the amount *OA*, and this expenditure volume will produce a sales level *OS* that is less than the maximum sales level shown by the advertising saturation level.

FIGURE 5-8

SALES-ADVERTISING MODEL



This, briefly, is the nature of the decision problem involved if management is to arrive at a scientific estimate of its advertising budget. More refined techniques and guides exist, however, which will be outlined in the following chapter. At the present we shall confine our attention to the practical aspects of translating these theoretical concepts into meaningful numerical estimates.

Measurement. In measuring short-run advertising effectiveness, a useful concept to employ is the *advertising elasticity of demand.* Like other elasticity notions, it is defined as the percentage change in sales (or market share) resulting from a 1 per cent change in advertising outlay. This elasticity coefficient may be affected by a number of factors such as: (1) the stage of the product's market development; (2) the extent to which competitors react to the company's advertising, either by further advertising or by increased merchandising efforts; (3) the quality and quantity of the company's past and present advertising relative to that of competitors', since variations in *qualitative* factors (e.g., choice of media) obscure the effects of *quantitative* variations in advertising outlay; (4) the importance of nonadvertising demand determinants such as growth trends, prices, incomes, etc., and the extent to which these can be averaged out in conducting the analysis; (5) the time interval that elapses between the advertising outlay and the sales response, which is difficult to predict because it depends on the type of product, advertisement, etc.; and (6) the influence of the "investment effect" of the company's past advertising and the extent to which this may be influential in affecting current and future sales as manifested by delayed and cumulative buying. In measuring advertising effectiveness, these are the sort of considerations that must be taken into account. Since the goal of the analysis is to discover what sales are as a result of advertising as compared to what they would have been without advertising, measurement methods must be devised that will allow and compensate for the above complexities. Several of these methods are outlined below.

Basically, the figures necessary for studying advertising effectiveness can be obtained either from historical data within the firm or from controlled experiments.[8] Historical data are often inadequate because they cover a period of time during which many unknown factors may have been significant in affecting sales, in addition to advertising as such. However, for companies whose market share is otherwise quite stable, such as with companies selling convenience goods which are largely income inelastic in demand, variations in advertising outlay over time may reveal significant sales differences so that a meaningful advertising-sales relation can be established and an elasticity coefficient computed. In actual measurement, multiple correlation must often be used in order to isolate advertising from other factors that may be responsible for sales. But correlation also requires that the significant causal factors affecting sales be measurable. Since this is not always the case, the use of correlation technique is not always possible. Controlled experiments, on the other hand, offer the opportunity for creating data, but the method may be relatively costly, particularly when the results must be subjected to further statistical (e.g., covariance) analysis. This consideration has prompted analysts to develop several alternative techniques for gathering data, four variations of which are outlined in the following paragraphs.

---

[8] A third approach is to base the analysis on the historical data of several firms. But this requires that products, prices, and other market characteristics be as similar as possible, so that differences in sales can be attributed to advertising outlay. Such uniformity is sufficiently rare, however, to make this approach quite impractical in most cases.

1. *Test Markets.* A common method of testing the effect of advertising copy, media, expenditure, or any combination is to select so-called "test markets" and then compare the results (e.g., first differences or percentage changes in sales) with other markets. Large sellers of food, for example, such as General Foods and General Mills, have used this approach extensively in planning their promotional strategies. Statistically, for consumer goods, the test market should be representative (usually in purchasing power and size of family) of the markets in which sales are ultimately intended. (Other criteria may include occupational characteristics, climate, religious and cultural factors, etc.) Also, the test markets should be sufficiently isolated from nearby nontest markets so as to minimize if not prevent the penetration of customers from the latter areas. The results derived from the test market experiments can then be blown up or projected on a national basis. Useful regression estimates can also be derived, depending on the ability to qauntify the advertising variable being tested.

2. *Split-Run Tests.* Another commonly employed method often used to test the qualitative aspects of advertising, e.g., copy, layout, illustrations, etc., is to design two different advertisements, each with a small coupon which the reader clips and mails in to receive a gift or further information. The different advertisements are inserted in different copies of a given newspaper or magazine at the same time and in the same place, thus eliminating the effects of these variables, and the resulting inquiries are then converted into a sales estimate. Keyed responses techniques along similar lines can also be used to measure the effectiveness of advertising outlay and in addition serve as a basis for deriving elasticity estimates. Statistically, it is essential to determine whether the change in sales is due to real causes or to random factors, before a decision can be made as to the significance of a particular causal factor.

3. *Sample Surveys.* Surveying a representative sample of readers can provide a rough indication of advertising effectiveness as well as useful data for further analysis. An illustration of the sort of estimates that can be made is given in Figure 5–9, showing the role of diminishing returns (increasing additional costs) in expanding the number of readers by adding successive publications. Although the total coverage curve

FIGURE 5-9

HOW THE LAW OF DIMINISHING RETURNS AFFECTS ADVERTISING MEDI

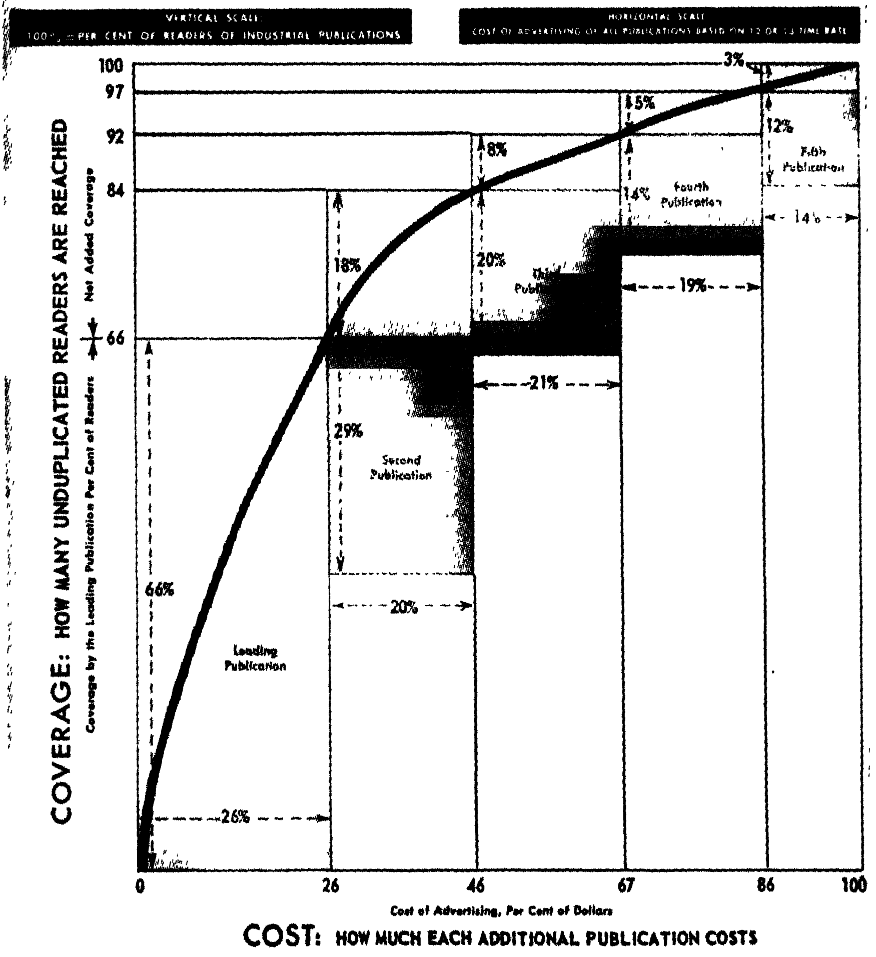
The chart shows how the law of diminishing returns operates when advertisers attempt to reach buying influences for their products who do not read the leading publication in a field.

The 100% figure in the chart is based on men reached by one or more of the first five magazines in nine different fields—more than 11,000 of them. The respondents are part of 42,878 who responded to the 1951 Cooperative Readership Study made over the customer and prospect lists of eighteen manufacturers.

The curve is an average one, and does not picture exactly the situation for any specific group of papers in a single field.

The chart demonstrates that it is un-economical, under average conditions, to select more than the one or two leading publications in the field to cover the important buying influences.

Note that fewer new readers are reached by each additional publication added, and that the cost of reaching these extra readers goes up.



Publication Ratings: 1st, 2nd, 3rd, 4th and 5th graded according to number of Industrial Readers who read each publication regularly

Source: *Laboratory of Advertising Performance,* McGraw-Hill Research, January, 1956, rev., p. 1120.1. Copyright 1955 by the McGraw-Hill Publishing Company, Inc. Reprinted by permission.

rises, the *additional* coverage compared to the *additional* cost (called "incremental coverage cost") soon becomes prohibitive. In a study made by Seymour Banks based on data provided by *Life* magazine studies covering the period 1938–41, a rising incremental selling cost was found in attempts to expand readership by means of additional magazines and additional issues.[9]

4. *Market Share.* Finally, a direct approach through market share is frequently used in either of at least two ways.

a) The firm may vary its own advertising outlay share as compared to that of competitors', and note the corresponding effect on sales in terms of market share. This is not always as difficult as it may seem at first blush. For industrial goods particularly, advertising media are usually well established (e.g., trade journals) and are used by all firms in the industry, so that management can keep fairly well abreast of approximate advertising outlays made by competitors. Further, several media trade associations publish data along these lines from which reasonably accurate estimates can be made.

b) Management may vary advertising outlay within narrow limits in order to discover the total outlay needed to retain market share. Advertising agencies often keep records of such data on a regional and national basis and maintain a running analysis as a service for their clients. Studies of this type have often revealed that when advertising outlay falls below a certain lower limit, market share declines sharply, and when the outlay exceeds a certain upper limit, the increase in sales is less than proportional to the increase in selling cost. This is often true with fairly standardized products (e.g., gasoline, toothpaste, soap, etc.) for which established brand preference is moderate at best and sales may be more influenced by factors other than advertising, and for products (especially industrial goods) that have a relatively stronger rational rather than emotional buying appeal.

Conclusion

This section has outlined the results of various empirical studies for the purpose of illustrating how quantitative estimates of demand can be of use in improving managerial decision making. The studies cited have been limited to those involving only simple relations, with price, income, substitutes, and advertising as the selected independent variables affecting sales. Emphasis has been placed on deriving elasticity estimates, since these are essential tools of scientific decision making and one of the major reasons for establishing quantitative relationships in business economics. In addition to the various elasticity measures discussed thus far, one other, that of *market-share elasticity* is worthy of mention. In this measure, the company's market share, which is the per cent of its sales as a portion of the industry's, becomes the dependent variable. The independent variable, however, may take one of several forms, such as average price differences or average price ratios between the company and the industry, or any other causal differences or ratios that may affect the firm's share of the market. The formulas, of course, would be of the same type as for the other elasticity measures already discussed. Taken together, these various measures provide a useful set of guides for implementing decisions and for establishing future plans.

---

[9] S. Banks, "The Use of Incremental Analysis in the Selection of Advertising Media," *Journal of Business* (October, 1946).

DEMAND FOR CONSUMER NONDURABLE GOODS

Against the background of simple correlation studies discussed in the previous section, we turn our attention in the remainder of this chapter to multiple correlation studies, or general demand functions of the form $Y = f(X_1, X_2, \mid X_3, X_4 \cdot \cdot \cdot X_n)$, where the terms to the left of the vertical bar indicate that now at least two factors are allowed to vary in arriving at an estimated sales function. Hence, the equation which states the relationship would be written for estimative purposes simply as $Y = f(X_1, X_2)$, since the fixed factors would not be included in the actual function. In this and the remaining sections of the chapter, a point raised earlier in Chapter 3 should be worth keeping in mind: as more variables are added to the equation, the gain in precision, or marginal accuracy, falls rapidly while the additional time and expense of the computation, or marginal cost, rises sharply. Forecasting perfection, therefore, may frequently have to be compromised with time and expense considerations in most kinds of practical work.

In the works to be discussed both here and in subsequent chapters, it should be noted that the studies were run by the mathematical rather than the graphic method of correlation, as the latter was explained in Chapter 3. This, however, should be no cause for concern, because the underlying principles of both methods are essentially the same. Therefore, even though some of the derived equations may appear quite complicated, the important thing to understand is not the actual equation itself, but the kind of variables selected and the interpretation of the resulting relationships. This requires a knowledge of economic theory primarily, and a basic familiarity with the concept of correlation as is at least conveyed by the graphic method.

Demand Determinants

A distinction between consumer durable and nondurable goods is often necessary in most types of consumer demand problems. The demand for nondurables i.e., perishables and semidurables, is frequently easier to measure because it involves current demand and is reflected in current market conditions. Usually, three classes of demand determinants enter into most empirical studies of nondurables, with each one modified as follows according to the specific nature of the product involved.

1. *Buying Power.* Disposable personal income, which is often taken as an indication of buying ability, is usually not an adequate measure of purchasing power and therefore must be adjusted accordingly. The adjustments take the form of adding: (*a*) cash stocks on hand; (*b*) consumers' credit, which usually shows sharp fluctuations that do not correspond with changes in disposable income; and (*c*) sometimes near-liquid assets, since these may also have a bearing on consumer optimism and willingness to buy. From this total is then deducted: (*a*) imputed income, of which one of the most important components is rent on owner-occupied homes; and (*b*) fixed outlay payments such as life insurance, pension funds, debt repayment, annuities, and necessary living costs in general. The result is a measure called *supernumerary* or *discretionary buying power*, which may be quite different from, and more significant than, disposable income for forecasting purposes.

2. *Demography.* This involves the population characteristics of the product concerned. For example, it may pertain to the number and characteristics of people in a study of the demand for food, or the number and characteristics of children in a study of the demand for toys, or the number and characteristics of automobiles in a study of the demand for tires. It is thus a recognition of the need for distinguishing between total market demand on the one hand, and *market segments* on the other, where the latter refers to the carving up of a total market into homogenous subgroups that have similar demand characteristics. Such segments may be derived in terms of income, social status, sex, age, educational level, geographic location, national origin, and numerous other dimensions as a means of arriving at more reliable estimates.

3. *Price.* Various direct and indirect influences of price may be employed, such as price differences and price ratios between the product concerned and competing or complementary products. In this way the relationships among the more relevant prices are brought into the predicting equation, thereby enhancing management control as well as forecasting accuracy.

By employing the appropriate forms of relevant variables, sales (S) can then be forecast by combining in an additive or multiplicative relationship the three factors buying power (B), demography (D), and price (P), as shown by the general formulas

$$S = B + D + P \quad \text{or} \quad S = BDP$$

and illustrated in the following case studies.

Demand for Gasoline[10]

What are the controlling forces affecting the demand for gasoline? Total sales of gasoline can be expressed as a function of three classes of factors: (1) the number of gasoline consuming units in use, which includes passenger cars, trucks, airplanes, buses, tractors, and other miscellaneous sources; (2) the average number of miles per unit, which in turn depends upon the composition of motor vehicles in use, changes in supernumerary income (i.e., disposable income less living costs), and gasoline prices; and (3) the average number of miles per gallon. A fourth variable, such as a time trend, is also needed as a measure of the influence of other factors not included in the above three. This embraces the average of all other miscellaneous influences that affect average mileage. The forecast equation is then given by the formula:

---

[10] The results of this study and of the two that follow, dealing with beer and with women's dresses, are summarized from a paper by L. D. Colburn of the Econo-

$$G = \{1.66T + [C + 5.5B][.458(3.15I)^{1.65}]4.74P^{-.52}\}17TT$$

where G = gasoline consumption in millions of barrels per calendar year
T = trucks in use July 1
C = cars in use July 1
B = buses in use July 1
I = supernumerary income
P = price of gasoline in cents per gallon (average per calendar year)
TT = time trend, which is a measure of those factors not otherwise included in the formula.

The equation reveals that there is a multiplicative relationship between the variables, the significance of each of which as demand-controlling factors may be outlined separately.

1. *Truck consumption* of gasoline, shown as the first term in the equation, was found to depend very little on both gasoline prices and general business conditions. The tendency on the part of truck operators has been to maintain a fairly constant mileage per truck by taking trucks out of operation when business declines. Therefore, the rate of truck consumption of gasoline is readily computed from the number of trucks in use, although average consumption has also increased with the size and type of truck. Diesel trucks and buses have grown somewhat in importance, but their combined production for the domestic market averages less than 2 per cent of the total.

2. *Passenger car and bus consumption,* as separate consuming units, revealed a narrowing spread between themselves because of the gain in the proportion of school buses in use, due in turn to two factors: the rapidly increasing number of school-age children, and the trend to suburban living. It was found in this study that gasoline consumption per passenger car (weighted relative to buses) tends to remain relatively constant in periods of economic stability, and that a distinct relationship be-

---

metric Institute, Inc., "Forecasting Sales of Semi-Durable and Perishable Goods" (1957). The Institute, under the late Dr. C. F. Roos's direction, pioneered in the application of econometrics to business forecasting, and the numerous studies made for its clients are at once probably more elaborate, durable, and sophisticated than those turned out by any consulting organization in the world.

tween the weighted average of cars and buses in use and gasoline consumption exists. This relationship, found by plotting fuel consumption (Y) against a weighted average of cars and buses in use (X) on a logarithmic scatter diagram, revealed a regression line whose equation was $Y = 17X$.

3. *Income and price variations* showed that a causal relationship existed with gasoline consumption per weighted passenger car. The use of supernumerary income as a measure of discretionary purchasing power is particularly useful in interpreting the results. Thus, peoples' working and living habits are such that a great deal of driving is necessary. Therefore, small variations in purchasing power exercise a relatively slight effect on gasoline consumption per car, so that there is a tendency for short-term fluctuations in gasoline consumption to be dampened. But with a substantial change in purchasing power as occurred in the early 'thirties, many persons were unable to operate their cars, and those that did operate them did not reduce their gasoline consumption proportionately. The result was that a sharp drop in purchasing power reduced the number of cars in operation, but the change in average gasoline consumption per car tended to be small. As a consequence, so long as supernumerary income has been in a generally rising trend as during the past decade, gasoline consumption for cars and buses could be reasonably well forecast without reference to this variable. But in periods of economic fluctuation, the inclusion of supernumerary income for prediction purposes turned out to be quite important as a means of improving forecasting accuracy. Mathematically, the relationship between gasoline consumption and the rate of change of supernumerary income, derived as explained in Chapter 3 by plotting the residuals from the previous regression line against the next independent variable (in this case supernumerary income) was found to be $Y = .458\ (3.15X)^{1.65}$.

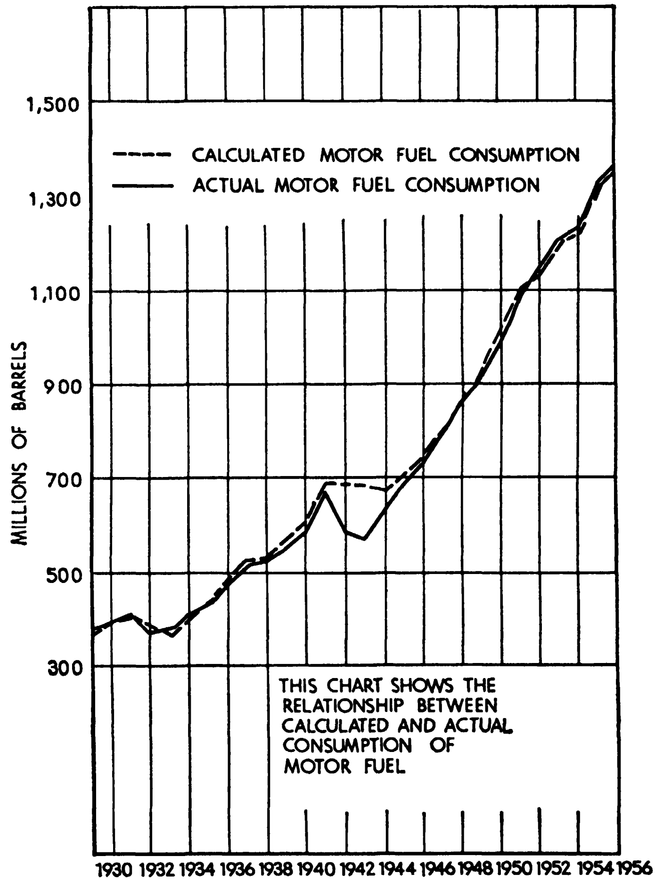
With respect to gasoline prices, the influence of these on sales proved to be not very large. The degree of elasticity varied somewhat in different regions of the country, with the greatest sensitivity to price changes in the low income regions as is to be expected. The mathematical formula for the relationship, representing the rate of change of gasoline consumption as a function of the rate of change in gasoline prices, derived from a scatter diagram plotted on logarithmic paper, turned out to be $Y = 4.74\ (X)^{-.52}$. In this case, unlike the previous ones, the regression line was negatively inclined. This is to be expected, since we know from economic theory that price exerts a negative influence on demand and that demand curves, therefore, slope downward from left to right.

4. *Time* is included in the formula as a catchall, representing the influence of all factors other than those stated above that may affect gasoline consumption. What are these "other" factors? They include the weight of vehicles, types of engines, the increasing use of diesel engines, highway mileage and maintenance, and trends in aviation and tractor gasoline requirements, to mention the most common ones. Being difficult to measure

or perhaps not necessary to measure in terms of the small gain in accuracy that would result, they are grouped together as residual factors and represented as a net residual or time trend which changes slowly with time. In a complete analysis, however, which is impractical and usually impossible, there would be no "residual" to be explained by a growth or time trend, because all relevant factors would be incorporated in the predicting equation. But as stated several times in earlier discussions, analysts are content to explain most of the variations in the dependent variable rather than all, so that the use of a time trend is both feasible and economical from a measurement and prediction standpoint.

FIGURE 5–10

Actual and Calculated Motor Fuel Consumption
(In Millions of Barrels)



Source: Econometric Institute, Inc.

Conclusion. What can be said as to the accuracy and durability of this analysis for forecasting purposes? An indication is revealed graphically in Figure 5–10. Published originally in 1942 and based on data beginning with the year 1929, the analysis has withstood the test of time most admirably. With the exception of the war years when gasoline was rationed, the broken line which is the graphic representation of the prediction equation presented above, corresponded very closely with the solid line representing actual gasoline consumption. It is unlikely that any other method of forecasting (see Chapter 2) could have produced results as accurate and as durable as has this econometric analysis of the demand for gasoline.

Demand for Beer

The general formula for consumers' nondurable goods discussed earlier can be employed to establish a good forecast of the sale of beer. The most important demand determinants may be measured in terms of three controlling variables.

1. *Real Disposable Income.* Disposable income is equal to personal income less all direct taxes, and real disposable income is simply the resulting figure deflated by the consumers' price index. The reason for this deflation has already been explained in the first section of this chapter.

2. *Population Segment.* It is believed in the beer industry, and supported by statistical evidence, that people in the 18 to 44 age group constitute the major purchasers of beer from civilian sources. Therefore, the per cent of the civilian population in this age group relative to the total population becomes another important variable affecting beer sales.

3. *Substitutes.* Wines and distilled spirits are rival products of beer. Hence their consumption represents the final variable needed to arrive at a forecast of beer sales, as will be seen shortly.

The beer industry uses beer withdrawals (in physical volume of barrels) as the major measure of beer sales. The problem, therefore, is to derive a relationship between beer withdrawals and the major independent variables stated above, which can then be used as a prediction formula. The nature of the relationship between beer withdrawals and each of these variables may be discussed briefly.

1. *Income.* Sales of beer, as measured by beer withdrawals, tend to vary directly with real disposable income when the data are plotted as a scatter diagram on logarithmic paper. Consumer credit is not used in the analysis as a supplement to income because beer sales are predominately cash sales and do not rest on the purchaser's decision to expand or contract his income commitments. The mathematical relationship, covering a period of twenty-three years, revealed that on the average the income sensitivity coefficient was 0.7, meaning that a 1 per cent change in real disposable income was accompanied by a 0.7 per cent change in beer sales. The formula for the sales-income relationship is $Y = 1.79X$. Although real disposable income has had a rising trend during much of this time, beer sales have not kept up with this full growth pattern and have actually declined relative to income since 1947. This has been due to the effects of the two remaining demand variables discussed next.

2. *Population.* There has been a decline in the percentage of people in the 18 to 44 year age group relative to the total population. This change in the age composition of the population has had an adverse effect on the sale of products marketed primarily to the young adult and middle-age group. With respect to beer, the nature of the relationship may be stated mathematically against the background of the graphic method of correlation explained in Chapter 3. As in the instances cited previously in this section, the method consists of taking the residuals about the regression line of the previous scatter diagram and plotting them against the next independent variable, in this case the per cent of total civilian population in the 18–44 age group. The equation for the resulting regression line, which expresses the relationship between beer sales and the rate of change of the relevant age group, is then found to be, in exponential form, $Y = 10^{(\,0066X-\,1250)}$, or in equivalent logarithmic form, $\log Y = .0066X + (9.8750 - 10)$. The latter equation, as explained in Chapter 3, is preferred because the calculations are more simplified.

3. *Substitutes.* Most of the remaining variation in beer sales (S) that existed after the previous two variables, income (I) and population (P), were included in the analysis, could be accounted for by including a third variable, consumption of wines and distilled spirits (C). Whereas beer sales tended to vary directly with real disposable income and with the relevant population segment, they tended to vary inversely with the increased consumption of substitutes. The mathematical relationship, found by taking the residuals in the S-P relationship and plotting them against C, was found to be $Y = 10^{(-\,020X+\,1080)}$, where Y is beer withdrawals and X is the consumption of wines and distilled spirits.

Conclusion. When the three separate demand factors are combined, they produce the result shown graphically in Figure 5–11, for which a time trend has been included in order to account for the remaining variables not treated separately as was done for income, population, and substitutes. The final formula, because of its more complicated nature, is presented and commented upon in the footnote below, where it may be overlooked if desired since the graphic result is sufficient for our purposes.[11]

---

[11] The predicting equation is:

$$S = 1.79\left(\frac{I}{\sigma I}\right)^{.7} 10^{\left[.0066\left(\frac{P}{\sigma P}\right)-.1250\right]+\left[-.020\left(\frac{C}{\sigma C}\right)+.1080\right]} (\sigma S)\ (TT)$$

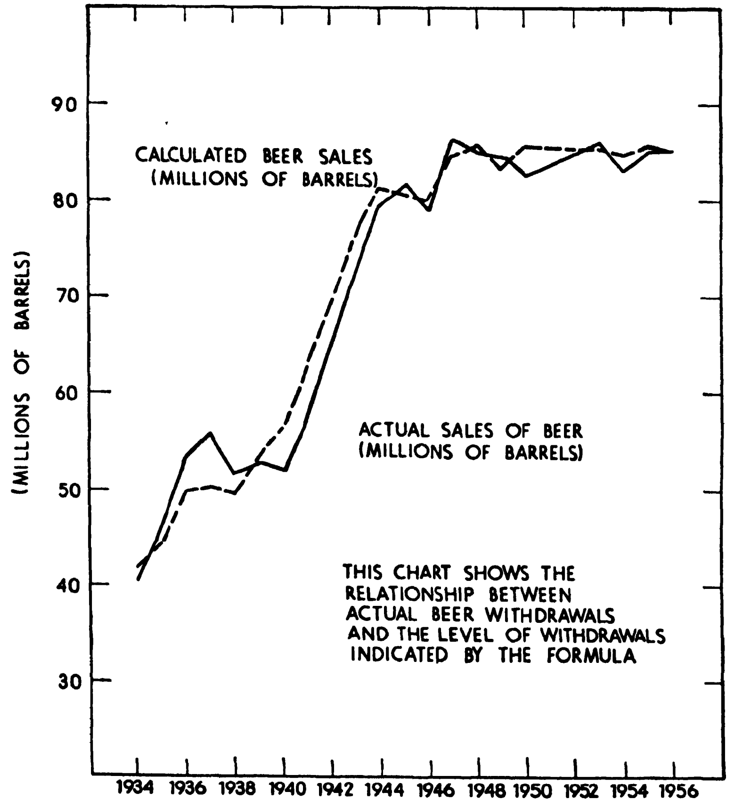
where S = beer sales, in millions of barrels annually; I = real disposable personal income, in billions of dollars per year (1947–49 prices); P = 18–44 age group as per cent of total population; C = consumption of wines and distilled spirits, in millions of gallons; and TT = time trend.

In the gasoline and beer studies, and in the following study on women's dresses, the regression lines between the factors were fitted freehand. Therefore, to obtain maximum likelihood solutions, each variate was expressed in units of its standard deviation (sigma, σ) and then plotted on equal scales. For a further explanation, see C. F. Roos, "A General Invariant Criterion of Fit for Lines and Planes Where All Variates are Subject to Error," *Metron* (Rome: February 28, 1937), p. 16.

It may be noted that the factors used to explain beer sales can be forecast with useful accuracy. Good predictions of disposable income are readily available from governmental and private sources, or can be prepared by the analyst himself; the 18 to 44 age group involves no difficulty in forecasting, except perhaps in time of war; and the cost of living, which changes slowly, usually moves in the direction indicated by the ratio of hourly earnings to output per man-hour. Both Figure 5–10 on gasoline sales and Figure 5–11 on beer sales provide a visual indication of how well the dependent variable could be predicted on the basis of forecasts of the relevant independent variables.

FIGURE 5–11

Actual and Calculated Beer Sales



Source: Econometric Institute, Inc.

Demand for Women's Dresses

A scientific forecast of the sale of women's dresses can be made by employing a few techniques slightly different from those encountered in the previous two studies.

Total sales of a broad product group such as substitutable outerwear (women's dresses, blouses, skirts, and sportswear) are related to external factors such as purchasing power and population characteristics. But the sales of individual apparel lines within the group, though they also respond to general economic and demographic changes, are a function of a third factor as well, namely style. A formula that attempts to forecast the sale of women's dresses, therefore, must involve the use of directly substitutable garments, particularly women's blouses, skirts, and sportswear. The approach taken in this study to arrive at a forecast of the sale of dresses was to forecast the index of combined sales of the outerwear group as a whole and then to evaluate the style trends separately.

Combined sales of outerwear are primarily a function of: (1) "consumer purchasing power" (i.e., a series, constructed by the Econometric Institute, Inc., which is composed of disposable income plus net changes in short- and intermediate-term consumer credit), and (2) a population factor. The population factor consists of (*a*) the ratio of the 14 to 54-year female age group to the total population, and (*b*) the ratio of females single, divorced, and widowed to total females over 14 years of age. When these determinants are correctly combined, they produce a calculated sales level around which actual sales have fluctuated within narrow limits.

The two population factors used in the analysis were characterized by a gradual, relatively smooth, declining trend for the period of the analysis, 1929–56. Their influence was therefore removed from the composite sales index by a statistical adjustment, made by calculating a combined population index (1947–49 = 100) with a weight of 1 assigned to the ratio of females 14 to 55 years of age to total population and a weight of 5 assigned to the ratio of females single, divorced, and widowed to total females over 14 years of age. The composite sales index, also on a 1947–48 base period, was then divided by the population index, in order to "reduce" the sales index as described in the first section of this chapter. The resulting series, which represents sales adjusted for population, was then charted against consumers' purchasing power as a scatter diagram on logarithmic paper. An upward regression line was revealed, the formula for which was $Y = .0856X^{1.1}$, where Y is apparel sales and X is consumers' purchasing power. Calculations based on the population factor and purchasing power were then divided into actual sales.

The final result for calculating sales of these apparel lines from 1929 to 1956 is shown in Figure 5–12, while Figure 5–13 compares the two selected indexes relative to the composite index as an indication of the changing pattern of substitutability. As Figure 5–12 shows,[12] actual sales

---

[12] The final formula for calculating sales of these apparel lines is:

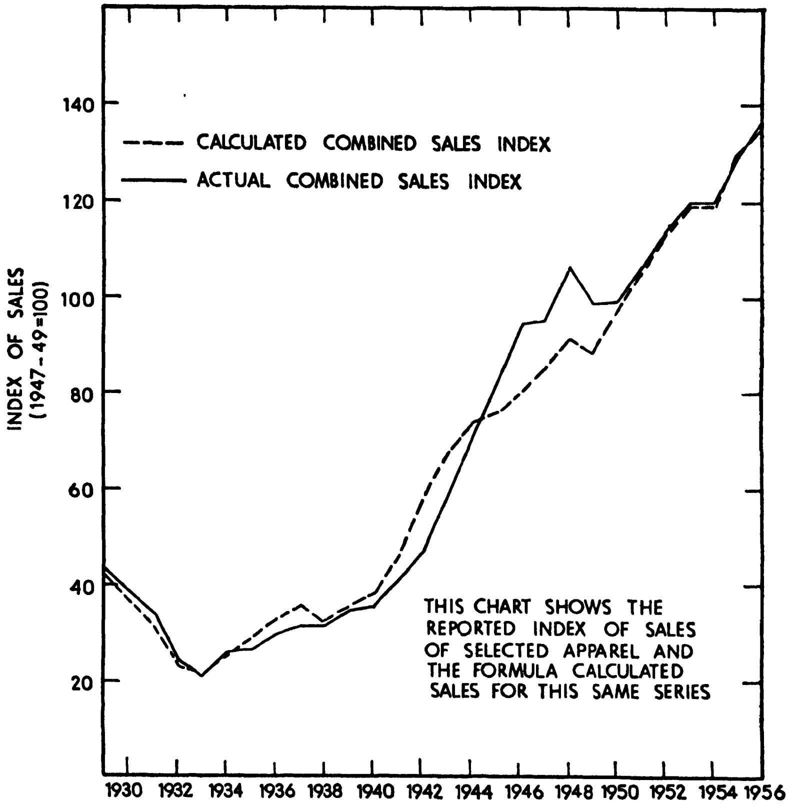
$$S = \left[\frac{\tfrac{1}{6}(P_1 + 5P_2)}{\sigma\tfrac{1}{6}(P_1 + 5P_2)}\right]\left[.0856\left(\frac{I}{\sigma I}\right)^{1.1}\right](\sigma S)$$

where $S =$ annual sales; $I =$ consumer purchasing power in billions of dollars per year; $P_1 =$ ratio of females 14 to 55 years of age to total population; and $P_2 =$ ratio of females single, divorced, and widowed to total females over 14 years of age.

corresponded closely to calculated sales from 1929 through 1940. During the war period actual sales fell below the calculated level due to merchandise shortages, while the reverse disparity occurred during the postwar adjustment period. By 1950, however, the adjustment was completed and actual sales have since fluctuated around the calculated level within a 2 per cent range.

FIGURE 5-12

ACTUAL AND CALCULATED DRESS SALES



Source: Econometric Institute, Inc.

Demand for Meat: Simultaneous-Equation Model

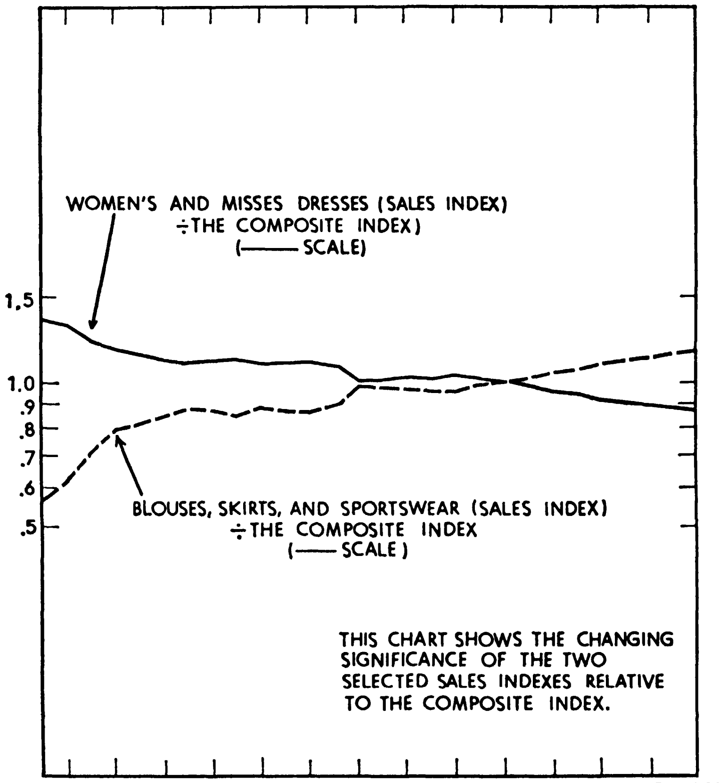
Whereas the previous studies employed the single-equation method of econometric forecasting, we may illustrate now the results of a simultaneous-equation model constructed by V. I. West in an analysis of the demand for meat.[13] It will be recalled from the discussion in Chapters 2 and 3 that a single-equation model might be preferred to a simultaneous-

[13] "An Analysis of the Demand for Meat Using a Simultaneous Estimation of Structural Equations," in E. J. Working, *Demand for Meat,* Appendix C, pp. 129-33.

equation model if the independent variables are predetermined, i.e., if they occur outside or prior to the given demand structure rather than inside of it. On the other hand, if the independent variables are not predetermined, as in many macroeconomic forecasts of the total economy, the controlling (independent) variables will be affected by factors within the given demand structure and hence a single-equation model would lead to biased estimates of the structural parameters or coefficients. In such cases, a better weighting of the variables might be achieved by a simultaneous-equation model instead.

FIGURE 5-13

Trends of Competing Sales Indexes of Women's Outerwear



Source: Econometric Institute, Inc.

West considered that meat prices, meat consumption, and income of any given year were inside or endogenous variables. These in turn were dependent upon the outside or exogenous variables: (1) production of meat, (2) investment expenditures, (3) income of the previous year, and (4) time. The last item, "time," is a catchall representing various unspecified factors which may exert a trend influence within the equations. It might be pointed out that production was regarded as an exogenous variable (with respect to this system or model) on the assumption that the amount of meat produced in a given year is not materially affected by the prevailing price of meat in that year.

The quantity of meat demanded was assumed to be influenced by retail prices, real income in the given year and previous year, and a time trend encompassing consumer desires and habits.[14] In order to express these relationships in equation form, the following X's were used to represent endogenous variables and Y's to designate exogenous variables:

X_1 = quantity of meat demanded per capita in a given year
X_2 = real (deflated) retail meat prices in the given year
X_3 = real (deflated) disposable income per capita in the given year
Y_4 = production of meat per capita
Y_5 = real (deflated) investment expenditure
Y_6 = time, annual (origin 1921)
Y_7 = real disposable income per capita of previous year $[X_3(t-1)]$
m = variables not specifically included in the demand equation but which affect demand.

The method of analysis used to estimate the parameters of the equation was based on maximum likelihood criteria. This is a less commonly employed method than the widely used least squares procedure, and it is recommended in some cases where certain variables in the system are jointly dependent (such as prices and income in this example). The following equations contain the estimates of the parameters obtained by those methods, based on the period 1921 to 1941:

$$X_1 = 157.46 - .91X_2 + .60X_3 - .52Y_6 + .07Y_7 + m_1 \qquad (1)$$
$$X_1 = 28.68 + .04X_2 + .74Y_4 + .09Y_6 + m_2 \qquad (2)$$
$$X_3 = 55.32 + .20Y_5 + .30Y_5 + .32Y_7 + m_3 \qquad (3)$$

The slope of the "demand curve" (change in consumption, ΔX_1, per unit change in price, ΔX_2) is $-.91$. The elasticity of demand at the point of averages is $-.63$.

The demand equation (1) was also fitted by the more traditional method of least squares in order to provide a comparison of results, particularly of elasticities. The single-equation model took the form:

$$X_1 = 156.07 - .88X_2 + .63X_3 - .54Y_6 + .03Y_7 \qquad (4)$$

The elasticity in this case is $-.61$. Clearly, the similarity between equations (4) and (1) is substantial, especially in the light of the uncertainties due to possible errors in choice of equation form. Accordingly, West concludes that the elasticity of demand for meat estimated from the systems model does not differ essentially from the elasticity estimate based on the single-equation model. From a practical standpoint, therefore, we may also conclude that, in view of the time and cost considerations relative to the added precision, the single-equation method might well have been preferable in this case. In any event, it is of interest to note that the results of this study proved to be of value to one of the major meatpacking companies in some of its planning operations.

---

[14] In addition, other factors were assumed to influence demand, but these were treated as being normally distributed and random in nature (i.e., stochastic).

DEMAND FOR CONSUMER DURABLE GOODS

A distinction between nondurable and durable goods is necessary for most types of demand problems because of the different sets of factors that characterize both classes of commodities. As in the previous section, therefore, we may begin by outlining the theoretical principles that serve as a guide for measurement, and then devote the remainder of the discussion to an examination of the results of several empirical studies.

Demand Determinants

How does the economic nature of the demand for consumer durable goods differ from that of consumer nondurables? In other words, can a special set of characteristics be established that will be useful for deriving quantitative relationships and for interpreting numerical results? The answer is given in the paragraphs below, in which are outlined some of the more important characteristics of the demand for consumer durable goods.

1. Consumer durables, precisely because they are durable, are not consumed in a single act as are foods, for instance, but instead dole out their services or are consumed over a period of time. This time-use characteristic raises a particular problem for forecasting purposes: because the good is durable, the consumer has to make a decision between (*a*) using the good longer by repairing it if necessary, or (*b*) disposing of it and replacing it with a new one. The significance of the former alternative was well illustrated during World War II, when the scrappage rate on automobiles and other hard-to-get durables dropped sharply. The choice of the latter alternative involves three further decisions: whether to scrap the good, trade it in, or sell it used. Whatever the choice, it may depend as much on noneconomic factors, such as one's social status and desire for prestige, as on economic factors, which include product obsolescence, income, and other conditions to be discussed below.

2. Durable goods usually require special facilities for their consumption. This use-facilities characteristic is exemplified by the need for roads and gas stations in the consumption of automobiles, wired homes for the consumption of electric refrigerators and washing machines, wired homes plus leisure time for the consumption of television sets, and dock facilities plus leisure time for the consumption of yachts. These use fa-

cilities, since they condition the sale of the good, must often be recognized in making a meaningful demand analysis and forecast.

3. Consumer durables are generally consumed by more than one person, as with a family consuming an automobile, refrigerator, or television set. The decision to purchase, therefore, may be influenced by family characteristics such as size of families and the age distributions of adults and children, as well as price, income, and other considerations. This is another form of the demography characteristic discussed in the previous section.

The above characteristics give rise to a particular type of market structure with respect to durable goods. Thus, the total demand for durables is really the sum of two demands: (1) a new-owner demand which serves to expand the existing stock of the good in consumers' inventories, and (2) a replacement demand which bears a particular relationship to both the existing stock of the good at any given time and to the size of the stock over a period of time. The significance of replacement demand should not be underestimated, as is a frequent tendency on the part of many managements that think solely in terms of finding new markets as a means of expanding sales. For long-established products, such as refrigerators and automobiles, the replacement market may represent as much as half, or well over half, the total market for the product. And even for less-established products, replacement demand may nevertheless constitute a substantial proportion of total sales in many instances.

The replacement demand for consumer durables tends to grow with consumers' stocks. For certain well-established products, especially automobiles, refrigerators, and radios, analysts have constructed life expectancy tables and survival curves which they apply to consumers' stocks in order to estimate average replacement rates. The automobile companies have derived such data at various times, based on scrappage rates, registration figures, and the like, a notable example of which appears in the Econometric Institute's celebrated study of automobile demand discussed later in this section. In general, it has been found that actual scrapping rates tend to depend mainly on purchasing power and production, and to a lesser extent on physical construction and operating costs. Thus, when purchasing power is increasing, scrapping rates tend to exceed theoretical expectations, and when purchasing power is declining, scrappage falls short of theoretical values. Also, when demand (sales) significantly exceeds production, scrappage is less than the theoretical value, but as production catches up with demand the scrappage rate tends to increase.

Basic Equation. In light of the above, the basic demand or sales equation for durable goods may be written

$$S = N + R$$

where S represents total sales of the good, N is new-owner demand or the increase in the existing stock, and R is replacement demand as measured by the scrappage of old units. Each of these independent variables may be forecast separately as explained below.

1. Replacement demand, or scrappage amounts, can be estimated by the use of life-expectancy tables or survival curves. If the required data are unavailable and cannot be estimated from published sources,[15] they can always be derived from sample surveys of consumers, or from consumer panel records which are maintained by many advertising agencies and marketing research firms. A well-constructed sample design can yield the needed scrappage data for virtually any consumer durable good. These data can then be correlated, usually with disposable income, number of households, and perhaps one or two other variables, to produce quite reliable and accurate forecasts of replacement rates. The results of such methods are illustrated later in this section.

2. New-owner demand, the remaining independent variable in the basic equation, represents the change in consumer stocks of the product with respect to a unit change in time. That is

$$\text{New-Owner Demand} = \frac{\text{Change in Consumer Stocks}}{\text{Change in Time}}.$$

Alternatively, the same relationship is expressed more conveniently in symbols. Letting Δy represent the change in consumer stocks, and Δt the change in time, then new-owner demand, N, is

$$N = \frac{\Delta y}{\Delta t}.$$

This, in turn, depends on other conditions which may be analyzed as follows.

At any given time, there exists in the economy a set of economic and cultural conditions which, in combination, determine an upper limit toward which consumers are continually adjusting their stock of a durable good. This upper limit may be defined as the *maximum* or *optimum ownership level.* It is "maximum" in the sense that it serves as an approximate demand ceiling for durable goods; it is "optimum" in that it is a level toward which the actual volume of consumer stocks is continually gravitating. Statistically, the maximum ownership level depends, for most consumer durables, on such factors as purchasing power (e.g., supernumerary income), number of families, and perhaps other factors, according to the product. Thus, for radios, it might also include some proportion of the number of automobiles, and for refrigerators the number of wired homes.[16]

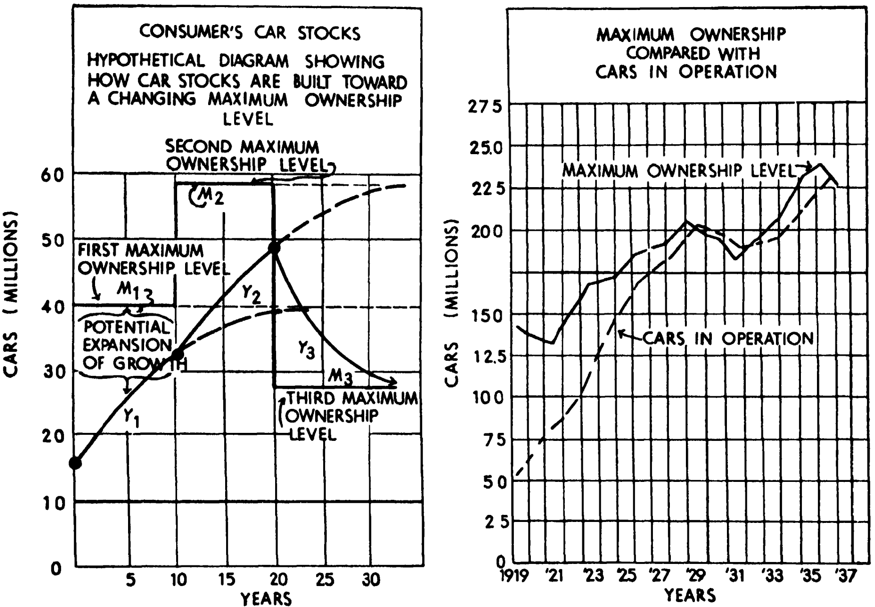
---

[15] An example of such a table that was used in the refreigerator study discussed below is B. F. Kimball, "A System of Life Tables for Physical Property Based on the Truncated Normal Distribution," *Econometrica* (October, 1947). Life-expectancy tables and survival curves have also been published by the automobile industry at various times, based on car registration figures, junk rates, etc. Kimball's tables are well worth examining as an illustration of the type of information needed, and from which the data can often be adjusted to fit particular circumstances as may be necessary.

The concept of maximum ownership level is portrayed graphically in Figure 5-14. In Chart A, there are three such levels, indicated by M_1, M_2, and M_3, each one being static until sufficient pressures are built up by its controlling independent variables to cause a change. Accordingly, for each M level, there is a corresponding y curve representing actual stocks in use, such as y_1, y_2, and y_3. Note that each of these stock curves tends to approach its corresponding maximum ownership level as shown by the dashed lines, but changes in the level cause shifts in the actual stock curve as well. It is evident, too, that the difference between M and y always represents the *potential expansion of growth* as shown on the chart. Finally, in Chart B, the actual empirical results for automobiles is given for the period 1919–38, indicating how the actual stock of cars is continually tending toward a changing maximum ownership level over time.

FIGURE 5-14

MAXIMUM OWNERSHIP LEVEL FOR AUTOMOBILES



A. HYPOTHETICAL

B ACTUAL

Source: Adapted from *Dynamics of Automobile Demand*, General Motors Corporation, 1938, p. 37.

---

[16] The concept was developed by C. F. Roos in his classic study of automobile demand discussed below. Also, in two talks delivered at an Econometric Institute seminar in 1957: C. F. Roos, "Techniques for Forecasting the Sale of Consumers' Durable Goods," and B. Slatin, "Case Studies in the Demand for Consumers' Durable Goods." Some results of the latter are summarized below.

The latter curve was derived from an equation in which the number of families, real supernumerary income per capita, and an index of replacement cost were the independent variables.

The rate of change of new-owner demand, N, which equals $\Delta y/\Delta t$, is dependent, therefore, on the potential expansion of growth. That is, the difference between the maximum ownership level and the actual stock of the product in use is $M - y$. But it is logical to suppose that this difference in itself is in some way proportional to the existing stock, y, already in use. Therefore, the equation for N may be written

$$N = \Delta y/\Delta t = ay(M - y)$$

where a represents a constant or parameter with respect to time, t, and it is the purpose of correlation analysis, as will be seen shortly, to determine this value of a in arriving at a demand equation.[17]

Demand for Automobiles

Of the numerous factors that may be influential in affecting the sale of automobiles, the following were selected, in a Commerce Department study,[18] as being of primary importance to have been incorporated into a demand equation: (1) income, (2) households, (3) price, and (4) average scrappage age.[19] By deriving a least-squares relationship between these factors and new-car sales for the period 1925–40, it was possible to explain most of the variations in car purchases in the prewar years, but not for the war and early postwar years because of the cessation of production and the backlog of demand. From 1949–51, however, an extension of the 1925–40 relationship showed that sales were once again becoming at least roughly consistent with prewar relationships (see Figure 5–15). Apparently, the demand function was quite durbale for forecasting purposes; it took the form:

$$Y = 0.0003X_1^{2\,5}X_2^{2\,3}X_3^{-1\,4}(0.932)^{X_4}$$

---

[17] Since M is assumed to be a continuous function of time, the equation may more appropriately be written $\Delta y/\Delta t = ay\,[M(t) - y]$, or: *rate of growth = potential expansion of growth*. If both sides of the equation are divided by y,

$$\frac{\Delta y/\Delta t}{y} = a[M(t) - y]$$

or: *percentage rate of growth = percentage expansion of growth.*

[18] See *Survey of Current Business* (May, 1950), (April, 1952). The last brings the previous two up to date for 1952. These studies covered the demand for automobiles, appliances, and furniture.

[19] Calculated from a least-squares regression for the years 1925–40, the equation is approximately $Y = 0.0001\ X_1^{1\ 1}X_2^{0\ 4}$, in which Y = total private passenger car registration in millions; X_1 = number of households in millions; X_2 = real disposable personal income in billions of 1939 dollars. The correlation is $R = 0.96$. The exponents are elasticities as will be explained shortly in the text.

in which

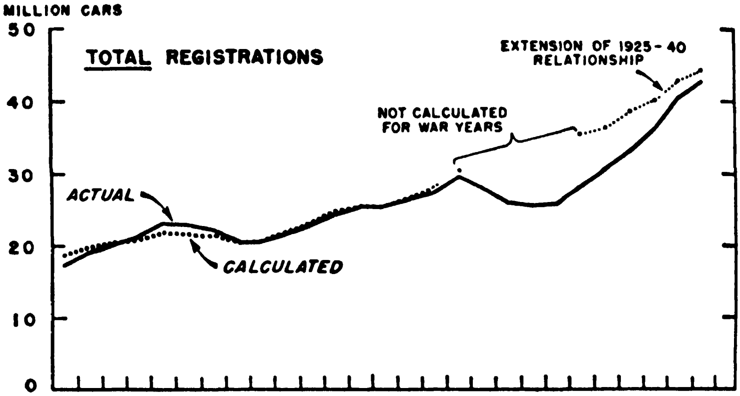
Y = new private passenger car registrations per 1,000 households
X_1 = real disposable income per household in 1939 dollars
X_2 = current annual real disposable income per household as a percentage of the preceding year in 1939 dollars
X_3 = percentage of average retail price of cars to consumers' prices, the latter measured by the Consumer Price Index
X_4 = average scrappage age.

The coefficient of multiple determination is $R^2 = 0.96$, which means, as noted in Chapter 3, that 96 per cent of the variations in the dependent variable are explained by the independent variables in the analysis. The results are shown graphically in Figure 5–15.

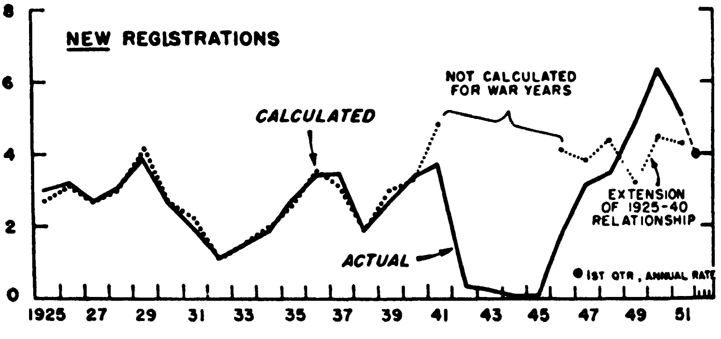
FIGURE 5–15

ACTUAL AND CALCULATED AUTOMOBILE SALES, 1925–51
Passenger Automobile Transportation Rebuilt Since the War

It took 6 years to bring autos in use about in line with income and population growth



Sales roughly consistent with prewar relationships



Source. *Survey of Current Business* (April, 1952).

The exponents in the equation are the elasticities. Thus, the most important factor affecting automobile sales is the real purchasing power of individuals as measured by real disposable income, X_1: excluding other factors, a 1 per cent change in real disposable income, X_1, is associated with a 2.5 per cent change in new-car sales, Y, during the period. Almost as important is the ratio of current to preceding year's income: a 1 per cent change in this variable, X_2, is associated with a 2.3 per cent change in sales, Y. Since these exponents are positive, the change in Y is in the same direction as the change in X; a negative exponent, such as for X_3, indicates that the changes occur in opposite directions.

A number of additional influences affecting automobile sales, such as trade-in allowances, credit conditions, operating costs, dealers' inventories, etc., could have been incorporated in the study, but the extra expense would have been substantially increased while the extra precision would have been relatively negligible, at least for the base period 1925–40. Perhaps the chief advantage of a more refined analysis would have been the durability of the function for forecasting purposes, which again depends upon balancing added costs against further accuracy. Thus, one of the best-known and most elaborate demand studies ever made was done by C. F. Roos and V. von Szeliski on the demand for automobiles, and is sometimes referred to as the General Motors study.[20] It probed deeply into the entire demand structure for automobiles taking into account the factors mentioned above as well as others. Roos and von Szeliski found that a separate analysis of new-owner and replacement sales would be impractical, mainly because there is some theoretical basis for doubting the complete independence of the two markets. Thus, if replacement sales are large, consumers will not add as many cars to their stocks as they would if few replacements were needed. If replacement requirements are small, some car owners may go on a two-car basis, or at least they may spend the money not needed for replacements in such a way that others can become new-car owners. Therefore, in the equation for total sales, new-owner sales and replacement sales were not expressed separately as such; however, a replacement pressure term was included in the formula on a par with the new-owners' term.

Combining the relevant factors, the problem was to fit the demand function

$$\text{Sales} = (\text{Income})\ (\text{Price})\ [\text{Car Stock}\ (\text{Maximum Ownership} - \text{Car Stock}) + \text{Replacement Pressure}]$$

to the data. Fitting the formula, they arrived at the equation

$$S = j^{1.2} p^{-.65} [.03C(M - C) + .65X]$$

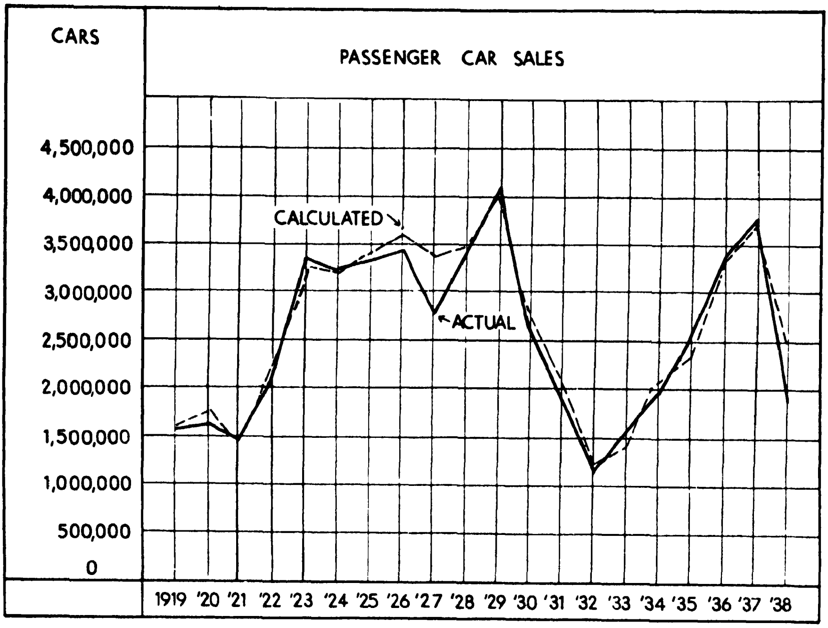
where S = new-car sales at retail, j = supernumerary income or income in excess of living costs, p = index of car prices, C = number of cars in use during the year, M = maximum ownership level, and X = replacement pressure, which was calculated by applying a shifting mortality table to the age distribution of passenger cars. (Note how this formula compares with the earlier discussion of the basic equation $S = N + R$ at the beginning of this section, and the formula for N involving the concept of maximum ownership level.) The exponents in the formula are the estimated elasticities and may be interpreted as in the Commerce Department's study discussed above.[21] The graphic results are shown in Figure 5-16.

---

[20] *The Dynamics of Automobile Demand* (General Motors Corp., 1938).

FIGURE 5-16

ACTUAL AND CALCULATED AUTOMOBILE SALES, 1919-38



Source: *Dynamics of Automobile Demand,* General Motors Corporation, 1938, p. 60

Demand for Appliances and Furniture

Figure 5-17 shows the results of a Commerce Department study of the demand for major appliances and furniture. The demand equation based on the years 1929-40 took the form

$$Y = -89.05 + 0.08X_1 - 0.35X_2$$

where Y = major household items per household in 1939 dollars, X_1 = real disposable income per household in 1939 dollars, and X_2 = time. In this study, 96 per cent of the variations in Y were explained by the independent factors in the equation; that is, $R^2 = 0.96$.

---

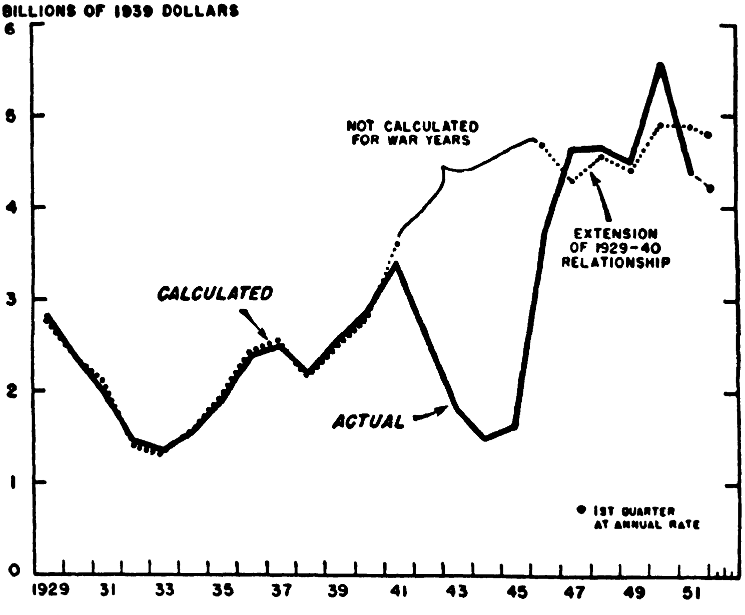
[21] Actually, there are statistical reasons for believing that the price elasticity was probably higher, perhaps 1.5 instead of .65 as shown, because of very few large swings in the price and because of its general downward trend for the period.

As in its automobile study, the Commerce Department's postwar projections of appliances and furniture demand held up fairly well using the above equation, as can be seen in Figure 5-17. The fact that the forecast or calculated estimates largely understated the actual results in the postwar period was due to the sudden growth trend in the sale of certain appliances, such as televisions, air conditioners, and freezers. The sales drop in 1951, with relatively few exceptions, was a follow-up to the wave of anticipatory buying that occurred in late 1950 and early 1951 (Korea) and then declined as inventories accumulated in factories and distribution channels.

FIGURE 5-17

ACTUAL AND CALCULATED MAJOR APPLIANCES AND FURNITURE SALES

Volume of Major Household Appliances and Furniture Purchases in 1951 Was Below the Prewar Relationship to Income



Source: U.S. Department of Commerce, Office of Business Economics.

This study was relatively simple and produced results that appeared to be quite adequate for many prediction problems that frequently confront business managers. It may compare, although briefly, with the results of two much more detailed, elaborate, and probably more durable studies made by the Econometric Institute—one for refrigerators and one for television sets. Although the procedure followed in these two studies was basically the same, there were some different analytical problems posed because refrigerators are a well-established product while television sets are still relatively new.[22]

Demand for Refrigerators. The procedure followed in analyzing the demand for refrigerators may be summarized as follows.

1. *Raw data.* The data available are on manufacturers' sales, expressed in physical units rather than on a dollar basis. These figures are not the same as for consumer sales, but over a period of years they would both be equal. However, manufacturers' sales probably show greater cyclical fluctuations than retail sales of the same commodity, because of inventory changes in the channels of distribution. Thus, when manufacturers' sales are plotted as a time series beginning with 1925, the cyclical changes appear to increase over time as the market approaches saturation. This is because the impact of sales due to increased acceptance of the product is diminishing, while changes in sales due to changes in consumer purchasing power are becoming more important.

2. *Refrigerators in Use.* The refrigerator stock was calculated from the Kimball life table mentioned earlier (footnote 15), based on a 25-year maximum life and the assumption that refrigerators were repaired rather than scrapped in the 1942–45 war period. According to the table, about 90 per cent of the refrigerators produced in a given year will still be in use ten years later. Therefore, the refrigerator inventory at the end of any given year can be obtained by simply summing the number still in use in that year among those produced in the earlier years. In this way a cumulative survival table may be constructed.

3. *Scrappage.* The indicated scrappage figure for each year was computed from the cumulative survival table derived in the previous step, by subtracting the inventory at the end of the previous year from the number of those units surviving to the end of the next year. For example, at the end of 1955 there were 42,104 thousand refrigerators in use, of which 40,114 thousand survived to the end of 1956. The former less the latter is the calculated or theoretical scrappage for 1956, and equals 1,990 thousand refrigerators.

4. *New-Owner Sales.* This figure represents the difference between total sales and scrappage. Over a period of years the actual and calculated new-owner sales would be equal, but in any given year the actual may be greater or less than the calculated figure. Thus, in this study, the actual new-owner sales plus the short-term or cyclical difference between actual and calculated scrappage were analyzed. This enabled the cyclical component of the scrapping figure to be included in the analysis, for as mentioned at the beginning of this section, scrapping tends to be above normal when income is rising, and below normal when income is falling. In short, the new-owner figure contains the same cyclical components as the total sales figure, being above its normal in periods of rising income and below its normal in periods of falling income. Thus, since actual

---

[22] B. Slatin, *op. cit.*, is the author of the two studies.

scrapping estimates were not available, no attempt at true separation of the data was made; nor, in fact, was such separation necessary.

5. *First Correlation: Percentage Change in Consumers' Stocks versus Consumers' Stocks.* As pointed out earlier with respect to the $S = N + R$ basic equation, the percentage change of consumers' stocks in use is directly proportional to the number in use (see also footnote 16). The first correlation problem, therefore, was to find the equation of the line that relates these two variables. This was done by plotting the annual percentage changes in consumers' stocks of refrigerators (representing demand for refrigerators) against the annual number in use, as calculated from the life table, on an ordinary scatter diagram. Two characteristics were immediately evident from the diagram: (*a*) the percentage changes in consumers' stocks had low points in 1932, 1938, and 1949, thus indicating that in addition to the annual stock in use, the percentage changes were also affected by general business conditions; and (*b*) the percentage changes in consumers' stocks showed wide variations in the early years, but these tended to diminish over time as saturation levels were approached. The line chosen to fit the data was one that averaged out these cyclical swings, rather than one that would average out all the points on the chart as in the least-squares method.

6. *Second Correlation: Maximum Ownership Level.* The next problem was to discover the factors most responsible for causing variations in the maximum or optimum ownership level—the level toward which consumers are continually adjusting their stock of refrigerators in use. The relevant controlling factors were found to be: (1) number of wired households, (2) supernumerary income, (3) net extension of installment credit, and (4) prices of household furnishings. These were combined into two variables, wired households on the one hand, and some function of income, credit extension, and price on the other. The intercorrelation that existed was removed and the two variables were combined in a multiplicative relationship. The result was to express the percentage of refrigerators in use adjusted for wired households, income, and prices.

7. *Final Results.* The constants and relationships derived from the previous steps were then combined to produce the predicting equation. The graphic results are shown in Figure 5–18 for the two periods, prewar and postwar,[23] while the percentage deviations of total sales from the calculated values in Figure 5–18 are shown in Figure 5–19.

---

[23] The equations were:

$$N_p = dy/dt = y\left\{H_u\left[.0078 + .0095\left(\frac{I + 3C}{.01P(10)^{.027T + 1.035}}\right)\right] - .00002Y\right\}$$

and

$$N = dy/dt = y\left\{H_w\left[.0045 + .011\left(\frac{I + 3C}{.01P(10)^{.027T + 1.035}}\right)\right] - .000016Y\right\}$$

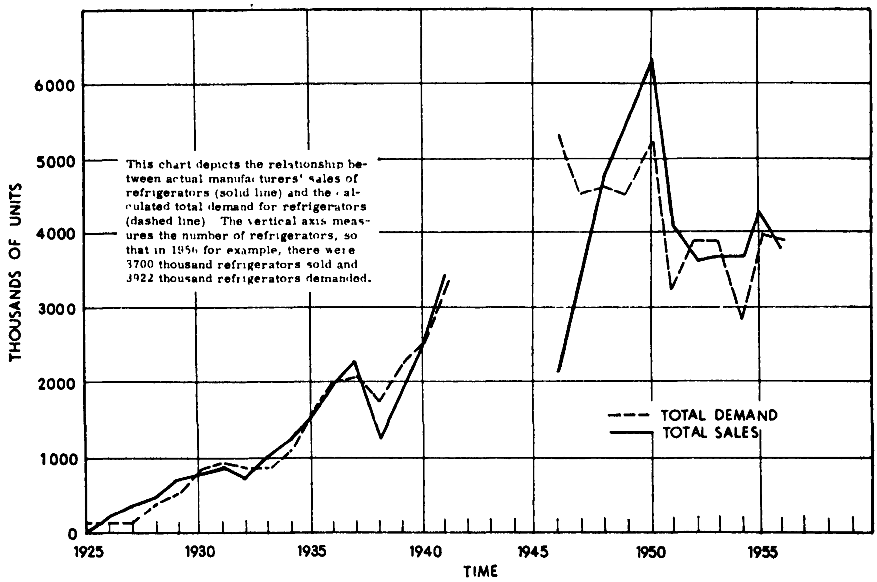
where N_p and N are prewar and postwar new-owner demand, respectively, and H_w = wired homes, I = supernumerary income, C = consumer credit, P = price index of

Interpretation. How may these charts be interpreted? Since it is their economic meaning rather than mathematical construction that is of primary use to management, let us see what conclusions can be drawn on the basis of the revealed relationships and our knowledge of economic behavior.

First, as to past results, there was a sharp cut in manufacturers' sales in 1938 below the calculated demand. This may have been due to a fear on the part of dealers of another depression following on the heels of the drastic 1931–33 experience. In 1950–51, on the other hand, actual sales were above the calculated level due to the buying scare engendered by the Korean war, while the consequent drop in actual sales in 1952–53 indicates the reaction to this previous overbuying. Finally, in 1953–54, purchasing power fell below its trend value, and this in turn caused the maximum ownership level to decline and thereby narrow the percentage growth from about 5 per cent down to 2 per cent.

FIGURE 5-18

REFRIGERATORS: MANUFACTURERS SALES AND DEMAND



This chart depicts the relationship between actual manufacturers' sales of refrigerators (solid line) and the calculated total demand for refrigerators (dashed line) The vertical axis measures the number of refrigerators, so that in 1956 for example, there were 3700 thousand refrigerators sold and 3922 thousand refrigerators demanded.

Source: Econometric Institute, Inc.

Second, as to the future, there are these factors to note. The refrigerator market is now fairly well saturated. "Saturation" may be said to exist when the percentage rate of growth of existing stock is small, or for

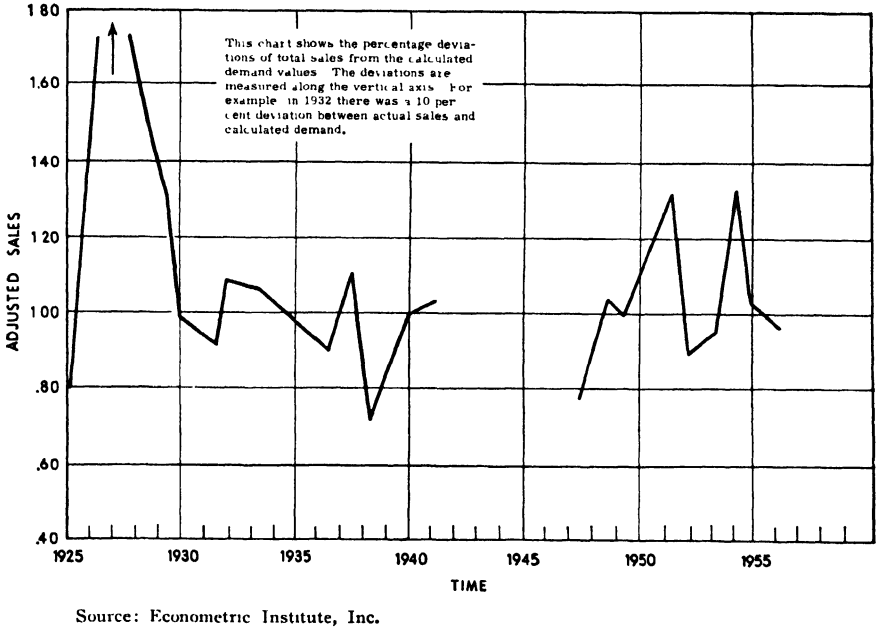
---

household furnishings, $T =$ given year means 1900, and y and t, as before, are stocks and time, respectively. Note how these equations compare with the Roos-von Szeliski automobile study, and how they typify, in a sense, the Econometric Institute's approach to demand forecasting.

practical purposes when it is about 4 per cent [i.e., $(\Delta y/\Delta t)/y = 4$ per cent]. This will occur as the existing stock, y, approaches its upper limit or maximum ownership level, M, so that potential new-owner demand, $M - y$, becomes small relative to y. Thus, suppose $M = 51$ million units and $y = 50$ million units. Then potential new-owner demand is 1 million units. If a small numerical increase occurs in one of the general variables such as purchasing power, credit terms, households, etc., M may rise to 52 million and thus increase potential new-owner demand from 1 million to 2 million units, or 100 per cent. But the *percentage rate of growth* of existing stock will only be $2 \div 50 = 4$ per cent. On the other hand, had the stock been 41 million instead of 50 million to begin with, the potential new-owner demand would have been 10 million instead of 1 million. Then an increase in M to 52 million would increase the potential new-owner demand from 10 million to 11 million units, or 10 per cent, but the *percentage rate of growth* of the stock would then be $11 \div 41 = 27$ per cent. In short, as the market approaches saturation, percentage changes in potential new-owner demand become larger, while the percentage rate of growth of existing stock becomes smaller as the y curve approaches its M ceiling at a decreasing rate (shown earlier in Figure 5–13a).

FIGURE 5–19

DEMAND FOR REFRIGERATORS: UNEXPLAINED VARIATION



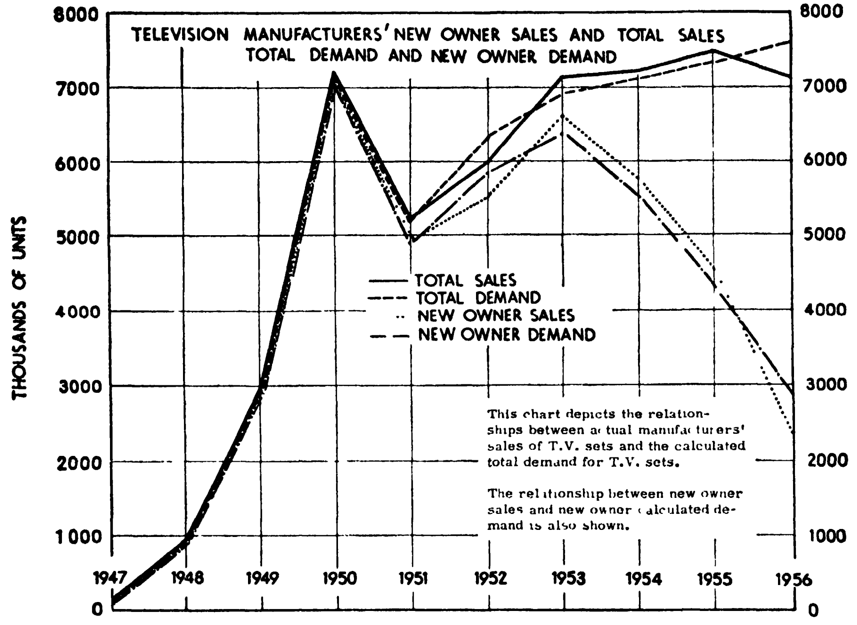
This chart shows the percentage deviations of total sales from the calculated demand values The deviations are measured along the vertical axis For example in 1932 there was a 10 per cent deviation between actual sales and calculated demand.

Source: Econometric Institute, Inc.

The significance of this for forecasting should be recognized. Since the refrigerator market is now fairly well saturated, forecast percentage increases should be checked against the percentage increases in new owners. As the existing stock (i.e., consumers' inventory) levels off and finally stabilizes itself as it now seems to be doing, further increases in refrigerator demand will proceed at about the same rate as the number of wired homes. In other words, once saturation is reached, forecasts of future sales can be based simply on changes in scrappage rates and on the additions to families. This assumes, however, that: (1) significant multiple ownership trends will not develop as has occurred, for example, in automobiles; (2) the growing market for home freezers will not reverse the trend

FIGURE 5–20

Actual and Calculated Sales of Television Sets



This chart depicts the relationships between actual manufacturers' sales of T.V. sets and the calculated total demand for T.V. sets.

The relationship between new owner sales and new owner calculated demand is also shown.

Source. Econometric Institute, Inc.

of freezer-refrigerator combinations and thereby affect the obsolescence rate on refrigerators; and (3) the present scrappage rate of 2 million units annually will increase gradually to 3 million in about five years, as the industry expects, so that new rates need not be re-estimated.

Demand for Television Sets. An analysis of the demand for television sets was conducted along the same lines as that for refrigerators. The specific problems encountered involved certain differences and similarities. As to the important differences: (1) television is a relatively new industry that made its significant commercial beginning in 1946; and (2) not until 1950 when 14-inch and 16-inch screen sets were priced to fit most budgets was television fully accepted and a definite break with past relationships established. As to the similarities with refrigerators: (1) available data required that the analysis be made in physical units rather than dollar sales; (2) manufacturers' rather than retail sales were used, thereby posing a wholesalers' and retailers' inventory problem; and (3) scrappage rates had to be estimated from data available on sets in use, on the basis of which a life table was constructed that revealed an average life of three years for sets produced in 1946–48, and an average life of 5½ years for sets produced after 1948.

The steps involved in arriving at the estimated relationships were basically the same as in the refrigerator study, and the final results are shown in Figure 5–20 for both new-owner sales, actual and calculated, and total sales, actual and calculated. As in the refrigerator study, the independent variables in the analysis were scrappage, stock (in this case of TV sets) in use, supernumerary income, net credit extension exclusive of automobiles, price index of house furnishings, and wired dwelling units.

From a forecasting standpoint, the problem is to predict both new-owner sales and replacement sales in the $S = N + R$ basic equation. New-owner demand, at the present number of sets per household, will probably be about 1 million sets per year in the early 1960's. Replacement sales, on the other hand, will depend upon the average life of sets. With a stock of 50 million sets in use in the early 'sixties, an average life of 5½ years as existed in 1957 would bring a scrappage rate of $50/5.5 = 9$ million sets per year, approximately, or about 18 per cent of 50 million. On the other hand, should the average life rise, say, to as much as seven years, the scrappage rate would build up to $50/7 = 7$ million sets annually, or about 14 per cent of 50 million. This indicates that the demand for television in the early 1960's will be in the range of 8 to 10 million sets per year, which is substantially above the level thus far achieved.

Conclusion

Classical economists, writing at a time when the analysis of demand was largely concerned with perishable and semidurable goods, focused on price as the most important variable affecting demand. But in the present-day economy, price is often insufficient as a controlling variable even where consumer nondurables are concerned. Other factors that must frequently be included are purchasing power and demographic characteristics if a reasonably good forecast of future sales is to be made.

For consumer durable goods, at least three factors—durability, price, and credit—should be recognized as conditions serving to complicate demand or sales forecasts. Unlike the sale of consumers' nondurables the consumption of which runs fairly parallel to purchase, the sale of consumers' durables does not exhibit this same pattern. Hence, a consumers' stock accumulates, which in itself is an important demand determinant, creating problems of life expectancy, replacement rates, etc., that must be forecast if successful predictions are to be made.

Price and credit conditions have interrelated effects. Since a large proportion of durable goods is sold on credit, finance and carrying charges should be viewed as a part of the price, as well as the real trade-in value of the good if it has any. Therefore, the use of price in a demand study should be employed in two ways: (1) The ratio of the price to the average life of the product should be considered—if the average life tends to change slowly, as is frequently the case, the principal effect will be to dampen the influence of price; (2) The significance of price in affecting down payment and carrying charges requires consideration—changes in credit terms can offset a price increase, for example, by lowering the down payment or by extending the payout period. The important thing for the analyst to recognize is that durability, price, and credit conditions have complex characteristics that must be understood if their effect on sales is to be measured and predicted.[24]

STRICTLY BUSINESS by McFeatters
Dale McFeatters
Reg. U.S. Pat. Office Publishers Syndicate

"I guess the firm upstairs is having its troubles!"

---

[24] C. F. Roos, *op. cit.*

Company Forecasts. The previous studies involving both durable and nondurable goods can serve as the chief guide in establishing long-run capacity requirements for the industries involved. From these forecasts, individual company forecasts can be made, usually by projecting the firm's typical market share into the forecast period. This assumes that the company will tend to bear the same relationship to the industry in the future that it has in the past. For short-run forecasting this assumption is often valid, but for long-range forecasting it must be recognized that fundamental changes may take place within the industry that could serve to change significantly the relationship between firms. If the latter occurs, a mere projection of market share will not produce a satisfactory forecast. Frequently, a scatter diagram comparing the company's output (on the vertical axis) with the industry's output (on the horizontal axis) will reveal whether the company tends to expand in or out of proportion to the industry during cyclical upswings and downswings. Information of this type can then serve as a guide to management in making decisions and formulating policies that will achieve the firm's desired future objective. Planned rather than haphazard growth is thus the keynote of effective coordination by management.

DEMAND FOR CAPITAL GOODS

A capital good, or producer's good, is a produced means of further production. These goods are not desired for themselves, but rather because they are used to produce consumer goods or services, or other capital goods which in turn will be used for the production of consumer goods or services. Examples of capital goods include machines, looms, tools, locomotives, ships, electronic computers, typewriters, and factory buildings. Strictly speaking, neat classifications are not always possible, for sometimes a capital good may also be a consumer good, depending upon its use (e.g., a jeep used on a farm for hauling purposes and as a means of transportation for pleasure purposes). What is important, however, is that the product becomes a capital good when it is purchased for the primary purpose of facilitating the production of another good to be sold for profit.

Demand Determinants

The approaches employed in analyzing the demand for capital goods as a whole differ substantially from those used in studying the demand for a specific capital good. The former problem area, that of the over-all demand for capital goods, has occupied the attention of business cycle theorists and econometricians for decades. Some of the typical questions are: What are the most important factors determining the over-all demand for capital goods? Do changes in the demand for consumers' goods cause changes in the demand for capital goods so that the latter may be said to be derived from the former? If the demand for capital goods really is a derived demand as economic theory suggests, what is the nature of its relationship with consumer goods demand and is this relationship the same in the upswing of a business cycle as it is in the downswing? Complete answers to these and similar questions have yet to be provided. However, some brief comments relating both to general and specific forecasts of capital goods demand will be helpful in pointing out a useful direction or way of thinking about problems of this type.

In 1955, C. F. Roos published the results of a study of producers' demand for durable equipment.[25] Demand, as measured by producers' expenditures, E, was related to corporate purchasing power, I (composed of the sum of retained corporate profits as measured by undistributed profits, new financing as measured by new capital issues, and capital consumption allowances as measured by depreciation and obsolescence); the BLS price index of manufactured goods, P_1; the BLS price index of metals and metal products, P_2; and the long-term interest rate, i, as measured by Moody's yield on high grade (Aaa) corporate bonds. For the period 1928–54, the predicting equation was

$$E = .012I(t - 3) - 74.49(P_1/P_2)_{t-6} + 2200$$

where t represents the forecast year and the subscript, $t - 6$, refers to the time interval six months earlier. The formula yielded an extremely good fit with actual producers' expenditures for the analysis period, with the exception of 1928 when corporations were using funds for call loans, World War II when actual expenditures were held below the desired demand due to wartime controls, and the postwar period 1946–47 when steel shortages were still acute. This formula was successfully used in the summer of 1956 to forecast: (1) a decline in domestic new orders for capital goods, (2) a decline in total capital goods production in January, 1957, and (3) a decline in industrial production as a whole in 1957;—all this at a time when opinion polls on producers' expenditures were optimistic, and most professional forecasters both in industry and government were predicting rising production in 1957.[26]

This analysis of the over-all demand for capital goods reveals certain essential points that are of interest from a forecasting standpoint (aside from the fact that it also challenges the opinions of those economists who hold that investment expenditures are largely unpredictable because they depend mainly on business sentiment and the psychology of investors).

[25] See his "Survey" article in *Econometrica* (October, 1955), especially pp. 391–94. For the background of the work leading up to this study and the earlier formulations of his estimating equations, see Roos and von Szeliski, "The Demand for Durable Goods," *Econometrica* (1943), pp. 97–122, and Roos, "The Demand for Investment Goods," *American Economic Review* (1948), pp. 311–20.

[26] C. F. Roos, "Techniques For Forecasting Sales of Capital Goods," (Econometric Institute seminar, 1957).

The fact is that in all industries, a relationship tends to exist between profits and capital outlays on the one hand, and the demand for the products of the industry on the other. When demand increases, profits rise and capital outlay expenditures relative to average outlay expenditures increase; when demand decreases, profits fall and capital outlay expenditures decrease relative to the average. Therefore, the operating rate—the ratio of production to capacity—of the user industries is an important variable to observe and to predict when forecasting the demand for capital goods, even at the industry level. Further, as the above study indicates, it may sometimes be possible to anticipate on an industry basis the expenditures on durable equipment by combining in some statistical manner a measure of corporate purchasing power (such as profits alone if the more accurate measure used above is not available), long-term interest rates, and the ratio of a consumers' goods price index to an index of machinery and equipment prices. The ratio of prices is a pressure index (see Chapter 2) and indicates the pressure of wage costs on the demand for newer and more efficient plant and equipment. Hence, a further indicator that may be of use is a measure of labor wage rates such as average hourly earnings, which, when combined with other relevant measures, may yield a further indication of management's desire to incur defensive investment in cost-saving capital equipment.

Demand for Steel

An application of econometric analysis to forecasting the sale of capital goods may be illustrated by two studies covering different periods of time for an important class of producers' goods—steel. The first study was made by R. H. Whitman; the second, by T. Yntema (now vice president of finance, Ford Motor Co.) for the U.S. Steel Corporation. Each is a significant contribution to a field in which relatively little has been done, probably because of the lack of adequate theoretical principles and statistical techniques until recent years.

Whitman Study. The pioneering work in the statistical analysis of demand for capital goods was done by R. H. Whitman and published in a remarkable article over twenty years ago.[27] He dealt with the demand for steel by considering various hypotheses that might be used to explain demand behavior. As an example he developed a dynamic demand function which involved a time derivative of the price and took the form for the period 1921–30 (using monthly data):

$$y = 1.49 - 1.27p + 6.27(dp/dt) + 4.641 - 0.03t$$

---

[27] R. H. Whitman, "The Statistical Law of Demand for a Producer's Good As Illustrated by the Demand for Steel," *Econometrica* (1936), pp. 138–52. Actually, Henry Moore attempted a statistical analysis of the demand for capital goods as early as 1914, but it was not as successful as his work on agricultural commodities. See his *Economic Cycles: Their Law and Cause.*

where

y = index of steel sales in millions of gross tons
p = price of steel in cents per pound, corrected for trend
dp/dt = rate of change of the price of steel over time (this is approximated by the first differences of p)
I = index of industrial production
t = time.

The multiple determination coefficient was $R^2 = 0.85$. All the regression coefficients except the last one are statistically significant, which means that there was probably no serious upward or downward shift in the demand for steel over the period.

How is the equation interpreted? The regression coefficient of p (-1.27) is negative: other things remaining the same, a rise in the price of steel brings a decline in its demand. The coefficient of the time derivative of the price of steel, dp/dt, is positive ($+6.27$) and substantially larger than the regression coefficient of p. This means that the demand for steel is speculative in that if steel prices rise, buyers anticipate still further increases and steel purchases expand; if steel prices go down, buyers expect further decreases and purchases decline.[28] Finally, the coefficient of the industrial production index is positive, indicating that the demand for steel bears a direct relation to general business conditions. These conclusions are in general agreement with economic theory regarding the demand for producers' goods, and Whitman's analysis provides a useful statistical verification of the theory.

Yntema Study. Almost as broad in scope as the GM study of automobile demand mentioned earlier, the U.S. Steel analysis of the demand for steel delved deeply into the economics of steel consumption and came up with a number of interesting and useful conclusions.[29] An essential purpose of the study was to discover the price elasticity of the demand for steel (as represented by bookings, shipping, or ingot production). Stated briefly, the analysis brought out the following facts concerning demand, based on conditions as they existed in the late 'thirties.

1. Steel is not a homogeneous product. The steel industry produces thousands of steel products, each with its own price, according to buyer specifications as to chemical, physical, shape, and dimension requirements. The industry's output is used in the production of many different kinds of products, each with separate output-determining characteristics. This creates an index number problem of combining the many products made from steel, the factors determining their outputs, and the quantity of steel used per unit of output into a reasonable number of economic composites.

---

[28] See, for example, C. F. Roos, *Dynamic Economics*, pp. 14 ff., for analyses of this type.

[29] "A Statistical Analysis of the Demand for Steel, 1913–1938," U.S. Steel Corporation, *T.N.E.C. Papers*, Vol. 1, pp. 169 ff. The study was under T. Yntema's supervision for presentation before the TNEC.

2. Steel is usually a raw material, not a finished good for consumption, and is used for making other products in such industries as transportation, construction, etc. Hence, the buyer's demand for steel is largely a derived demand and depends on conditions affecting the production and sale of products made from steel. These products are mainly durable goods and hence (unlike perishables) may be consumed at rates quite independent of their rates of production. The sale of steel thus being derived largely from the production of new durable goods exhibits wide fluctuations due to the "accelerator effect." Also, numerous nonprice factors exert a strong affect on the sale of steel, and this complicates price-quantity relations.

3. Steel can be stored, and hence changes in the size of steel inventories held by customers will be an important factor affecting the demand for steel. The size of such inventories will depend upon such conditions as buyers' anticipations, delivery times, cost of carrying inventories, etc. These factors give rise to an inventory range, bounded by an upper and lower limit within which customers' optimum inventory level is located.

4. Finally, steel is sold at different prices in different markets. There is no single price of steel throughout the entire economy. Instead price differentials are paid by different buyers for the same type of steel, and these differentials are of two main types: (a) more or less permanent differentials in different geographic areas due to cost structures and institutional arrangements in the pricing of steel (i.e., basing point system); (b) secret price concessions granted by sellers to buyers in order to attract sales away from competitors. The existence of these price differentials poses another index number problem of combining different prices into single composite prices for various kinds of steel. Ideally a separate demand analysis would be required for each group of buyers subject to the same price differentials.

After analyzing several general hypotheses as to the actual variables to be included, four mathematical demand relations were formulated for further examination by statistical techniques. Steel sales were represented by bookings (one equation), shipments (one equation), and ingot production (two equations). The independent variables were steel prices, industrial production, industrial profits, supernumerary income, rate of change of industrial production, and a time trend. Once the numerical values of the constants were known, it was possible to determine the change in the quantity of steel sold (as represented by bookings, shipments, or ingot production) associated with a given change in the independent variables. The price elasticity of demand for steel (with other demand factors held constant at their average value) was also determined for each of the four demand relations, the estimates of which are given in Table 5-2.

The elasticity estimates show that demand is quite inelastic: 1 per cent change in the price of steel is associated with a less than 1 per cent change in sales. It may also be noted that the difference in elasticity values between III and IV is due only to the difference between the estimates in steel sales used in each case, since the relations are identical in other respects. A conclusion drawn by the authors on the basis of this study was that since the demand for steel is so price inelastic, there is no sound basis for the frequently expressed view that the price of steel is a practical medium for stabilizing production. In other words, if the level of steel sales is to be ironed out, the problem must be attacked by controlling or manipulating factors other than price. This is quite the opposite of the traditional view held by classical economic theorists (concerning industries in pure competition) where price is held as the all-important vehicle for controlling production.

TABLE 5-2

DEMAND ELASTICITIES FOR STEEL

Steel Sales Measured by	*Demand Elasticity*
I. Production of ingots and castings*	+0.12
II. Production of ingots and castings†	+0.52
III. Steel shipments‡	−0.21
IV. Steel bookings§	−0.88

Independent variables are:
* Industrial production, time.
† Industrial production, time, rate of change of industrial production.
‡ Industrial profits, supernumerary income, time.
§ Industrial profits, supernumerary income, time.

Source: U S Steel Corporation, *T.N.E C Papers*, Vol I.

Demand for Portland Cement[30]

As stated earlier, the demand for capital goods may be approached from the standpoint of end-use analysis. The procedure consists of determining the most important users of the product in question, the factors that determine usage, and then fitting the results into a predicting equation. An illustration with respect to the sales of portland cement is typical.

The sale of portland cement depends upon: (1) the level of construction activity, and (2) the competitive position of cement relative to structural steel, masonry, and other substitutes. In this study, the construction industry, which actually embraces about a dozen markets (e.g., residential construction, industrial construction, highway construction, military installations, dams, airports, etc.) was conveniently grouped into three broad categories: producers' plant, *p;* highway, *h;* and all other construction, *a.* Each of these markets was then weighted according to its relative contribution to total consumption, based on Commerce Department published estimates of cement use by major sectors of the construction industry for the period 1947–49. According to these estimates, the allocation of total cement produced during that period was: $p = 19.5$ per cent, $h = 16.5$ per cent, and $a = 64$ per cent.

---

[30] Based on a talk delivered by M. C. Romano on the demand for construction materials (Econometric Institute Seminar, 1957).

To construct an end-use index, i.e., a measure of activity in end-use markets, the following procedure was employed. A separate activity index for p, for h, and for a, was calculated from Commerce Department data, adjusted for price changes, and placed on a 1947–49 = 100 base. Each activity index was then weighted by its percentage contribution as stated above, and the three were combined into a total market index. In other words, the formula for constructing the end-use index, E, was

$$E = .195P + .165H + .640A$$

where P, H, and A are indexes of producers' plant, highway, and "all other" construction, respectively, and the coefficients are the weights according to the percentage of total cement allocated to p, h, and a in 1947–49.

The series "portland cement shipments" was chosen as a measure of demand, D. Both sets of data, D and E, were then divided by their respective standard deviations in order to place the data on comparable scales. D, expressed in standard deviations, was next plotted against E, also expressed in standard deviations, on an ordinary scatter diagram. The relationship obtained was an upward-sloping regression line at a 45-degree angle to the base, indicating that the D-E relationship is unit elastic, i.e., a 1 per cent change in the end-use index is accompanied by a 1 per cent change in shipments of portland cement, other things remaining the same.

The remaining steps in the analysis involved the determination of the effect of other factors influencing shipments. These were lumped together as "unexplained variations" and treated as a net growth trend. The predicting equation, for the period 1920–56, turned out to be

$$D = \frac{E}{S_E} f(t) S_S$$

where

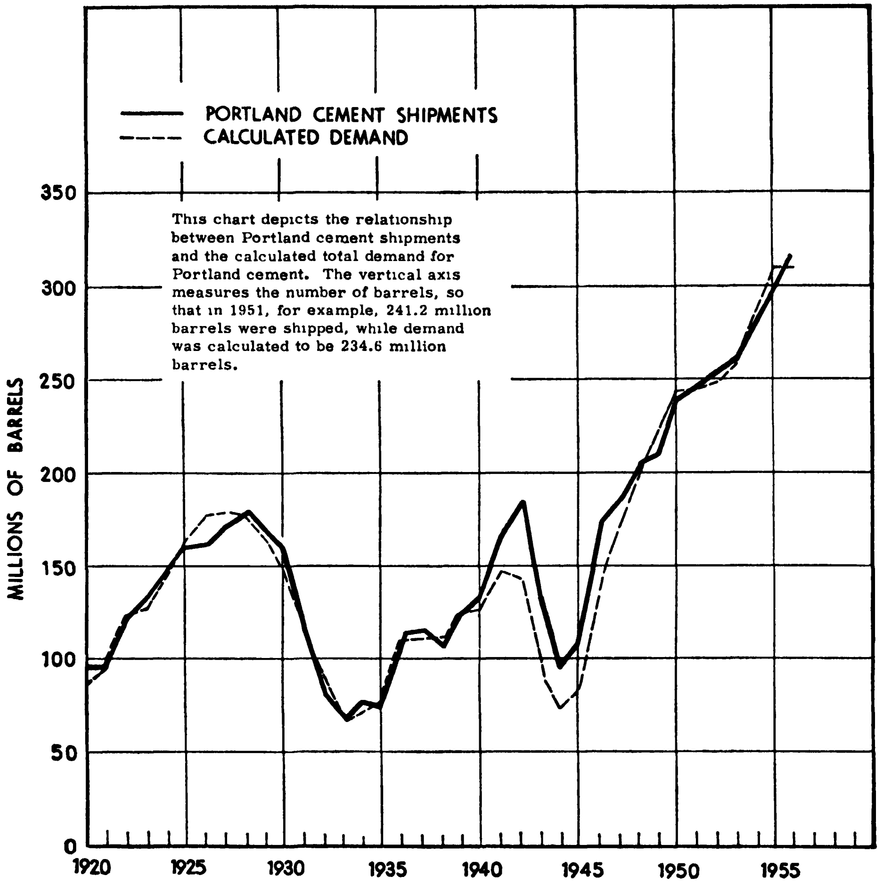
D = demand for portland cement
E = end-use index
S_E = standard deviation of the end-use index
S_S = standard deviation of portland cement shipments
$f(t)$ = net growth trend.

The graphic result is shown in Figure 5–21, while Figure 5–22 shows the percentage deviations between actual shipments and calculated demand.

Forecasts of demand for portland cement can be made, on an annual basis, by forecasting construction activity in the end-use markets and by projecting the net growth factor—and on a quarterly or monthly basis by including the normal seasonal variations. In forecasting, the probable change in inventory demand must also be estimated. This is sometimes quite difficult, but certain variables to be observed as a guide in making

FIGURE 5-21

PORTLAND CEMENT DEMAND: ACTUAL SHIPMENTS AND CALCULATED DEMAND



This chart depicts the relationship between Portland cement shipments and the calculated total demand for Portland cement. The vertical axis measures the number of barrels, so that in 1951, for example, 241.2 million barrels were shipped, while demand was calculated to be 234.6 million barrels.

Source: Econometric Institute, Inc.

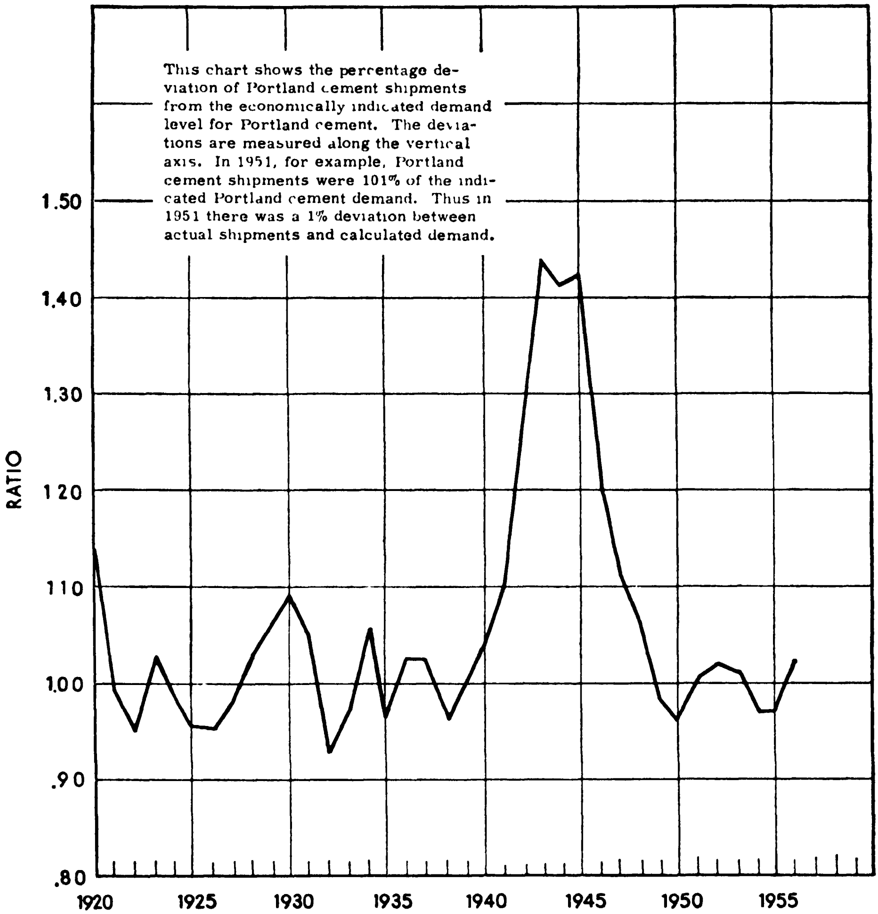
such predictions include contract awards, monetary changes, labor demands and their possible effects on cement prices, and changes in productive capacity. In other words, a close contact with production and market conditions is necessary if correct predictions are to be made.

Conclusion

The problems encountered in analyzing the demand for capital goods may be quite different from those involving consumer goods. The various products that may be classed as capital goods are subject to the complexities of derived demand, diversity of uses, and varying life expectancies. The first two factors are usually stressed in the literature, while the significance of the last is all too frequently overlooked. For instance, some capital goods are consumed in a single act and thus have physical lives that are

FIGURE 5-22

Portland Cement Demand: Unexplained Variation


This chart shows the percentage deviation of Portland cement shipments from the economically indicated demand level for Portland cement. The deviations are measured along the vertical axis. In 1951, for example, Portland cement shipments were 101% of the indicated Portland cement demand. Thus in 1951 there was a 1% deviation between actual shipments and calculated demand.

Source: Econometric Institute, Inc.

as short as the life of most consumer perishables. Industrial abrasives, soap, and certain chemicals are illustrative examples. Other capital goods have long physical lives and hence are capable of doling out their services for many years. However, their useful lives are curtailed due either to technological obsolescence, as with diesels replacing steam locomotives and jets replacing propeller-driven airplanes, or style obsolescence, as in the automobile industry where new bodies may require new tools and dies. Finally, still other capital goods, with appropriate modifications, repair, and upkeep, can have their useful lives prolonged for many years. Each particular type of capital good poses a different kind of problem and may require a particular kind of analysis. This is less frequently the case with consumer goods for which general categories and patterns of behavior can be more readily established.

BIBLIOGRAPHICAL NOTE

In addition to the classic work by Henry Schultz, *The Theory and Measurement of Demand*, which, though it is confined to agriculture, occupied a significant role in stimulating demand measurement for manufactured goods, the following briefer works are worth consulting both for theory and technique: H. R. Prest, "Some Experiments in Demand Analysis," *Review of Economics and Statistics* (1949); R. Stone, "The Analysis of Market Demand," *Journal of the Royal Statistical Society* (1945); and J. Marschak, "Economic Interdependence and Statistical Analysis," *Studies in Mathematical Economics and Econometrics*, pp. 135–50. All three of these require a background of at least a year or so of statistical analysis. On the other hand, further demand studies for specific commodities, and less mathematical in their exposition, are: J. Derksen and Rombouts, "The Demand for Bicycles in the Netherlands," *Econometrica* (1937); and V. von Szeliski and Paradiso, "Demand for Shoes as Affected by Price Levels and National Income," *ibid.*, (1936). Further studies, of course, include the works mentioned in the footnotes. A critical evaluation of statistical demand curves from the standpoint of economic theory is G. Stigler, "The Limitations of Statistical Demand Curves," *Journal of the American Statistical Association* (1939), pp. 464 ff.

Useful for more concise supplementary reading are discussions of demand measurement in Colberg, *et al.*, pp. 118–30, and Dean, pp. 187–246. More specialized but still not too technical for most students are the following. On single- and multiple-equation methods: E. J. Working, *Demand for Meat*, chap. 2; and K. A. Fox, *The Analysis of Demand for Farm Products*, pp. 6–37. Both are small booklets and, despite their titles, are of interest for the study of nonfood as well as food products. On advertising effectiveness: Alderson and Sessions, *Cost and Profit Outlook* (February, 1958); and an application of graphic correlation in S. Hollander, "A Rationale for Advertising Effectiveness," *Harvard Business Review* (January, 1949). Finally: Roos and Von Szeliski, "The Demand for Durable Goods," *Econometrica* (1943), pp. 97–122; Roos, "The Demand for Investment Goods," *American Economic Review* (1948), pp. 311–20; and Roos' "Survey . . . ," *Econometrica* (1955), especially pp. 388–95, provide a combination of sources that are basically nonmathematical and are particularly recommended for supplementary reading, as is the *Journal of Business* (January, 1954), which is devoted entirely to forecasting and contains a number of interesting and timely articles on various subjects.

QUESTIONS

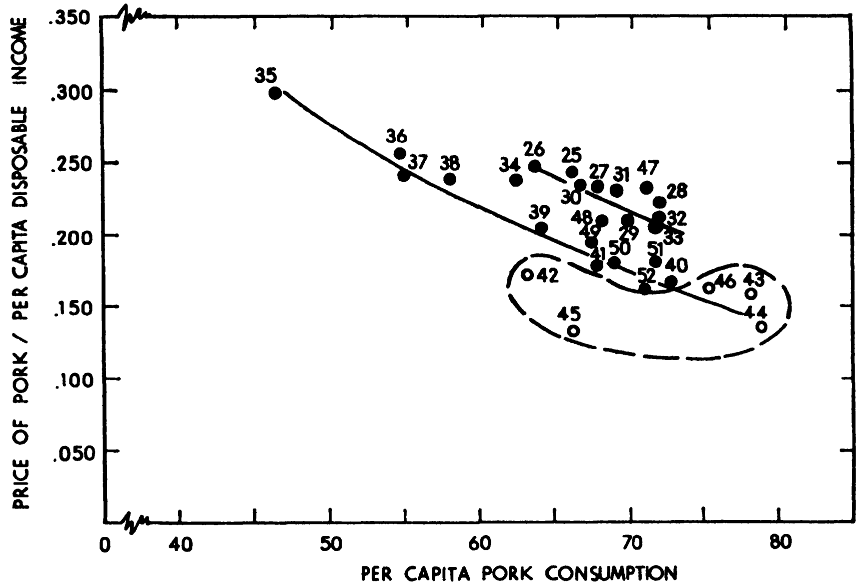
1. (*a*) What is the strict meaning of *demand* in economics? (*b*) Is this meaning a very realistic one from the businessman's standpoint? Explain.
2. In what different ways can a demand function be expressed?
3. In elementary economics, emphasis is placed on distinguishing between a "change in the quantity demanded" and a "change in demand." What are the meanings of these expressions and what are their implications from the econometric, i.e., measurement standpoint?
4. Distinguish between the historical or time-series method and the cross-sectional method of demand measurement.

5. Explain the nature of the statistical "reductions" or adjustments that must frequently be considered in the preliminary stages of a demand analysis.
6. What are the advantages and disadvantages to be considered in the choice of a simple versus multiple correlation model?
7. In general, what rule of thumb can you suggest as a guide for the inclusion of more controlling variables in a prediction equation? Explain.
8. (*a*) Formulate a general definition of elasticity. (*b*) Of what practical value is this concept?
9. Examine the accompanying diagram, Figure 5-23, of the demand for pork, 1925-52. (*a*) Has the demand for pork risen or fallen, and by approximately how much, percentagewise? (*b*) Without knowing anything as to the reasons for the shift, interpret the meaning from the graph alone. (*c*) In what year did the shift in demand actually occur? (*d*) Can you account for why the shift occurred? (Hint: Refer back to Figure 5-1, p. 141.) (*e*) What is the significance of the encircled dots? What is the significance of 1942 and 1945 compared to these encircled dots? What is the significance of 1942 and 1945 compared to these other years? (*f*) Estimate the more recent demand elasticity for pork and interpret your result. Is this elasticity greater or less than for beef?

FIGURE 5-23

DEMAND FOR PORK

Price of Pork Divided by Per Capita Income, Plotted against Per Capita Pork Consumption, Annually, 1925-52



Source: G. S. Shepherd, *Marketing Farm Products*, 3d ed., p. 311. Reprinted by permission of the Iowa State College Press

10. (*a*) Distinguish between the slope of a curve and its elasticity. (*b*) How may slope and elasticity be made comparable?

11. Refer to the department store study of a millinery product, p. 143. What-economic "dangers" are inherent in this type of study?
12. (*a*) What is the "consumption function"? (*b*) To what extent is total consumption predictable, as evidenced by empirical studies? (*c*) Which is typically easier to forecast: total consumption expenditures, or expenditures on a class of consumer goods, e.g., clothing or automobiles? Why?
13. (*a*) Derive a formula that can be used to measure the income sensitivity of consumer expenditures. (Hint: Define the term first, and, as a guide, refer back to the formula for price elasticity of demand expressed earlier. (*b*) If you derived a coefficient of 2.8 for a particular product group, how would you interpret this value for a 1 per cent change in disposable income? For a 10 per cent change?
14. What practical uses are there to deriving sales-income relations on a stage and regional level?
15. In your own words, distinguish between the cross elasticity of demand and the elasticity of substitution, without defining them as in the text.
16. (*a*) What basic decisions are involved in advertising budgeting? (*b*) The optimum advertising budget is not the one that maximizes sales. Why? Explain in terms of the sales-advertising model of Figure 5–8.
17. (*a*) Derive a formula for the measurement of the advertising elasticity of demand. (*b*) What is the primary use of this measure? (*c*) Where are the necessary data usually obtained for such measures? (*d*) What types of methods are typically employed to gather advertising data?
18. What underlying principle, if any, do you see in operation both in Figure 5–8, p. 156, and in Figure 5–9, p. 159? Explain.
19. In general, of what value are elasticity measures to management?
20. (*a*) What controlling factors are usually involved in affecting the demand for consumer nondurables? (*b*) Define and describe the concept of supernumerary or discretionary income. How does it compare with disposable income? (*c*) What is meant by "demography" as a controlling factor? (*d*) How is price used as a controlling variable in forecasting and demand measurement? (*e*) State the basic equation for forecasting consumer nondurable goods sales.
21. How, in terms of concept and technique, did the studies of the demand for gasoline and for beer differ from the study of women's outerwear? Explain.
22. What was the reason for employing a simultaneous- rather than single-equation model in the meat study?
23. What economic characteristics are peculiar to most consumer durables and which tend to make the problem of forecasting their sales more complicated?
24. (*a*) State the basic demand equation for consumer durables. Explain. (*b*) What is the nature of the relationship between scrappage rates and income? (*c*) How are scrappage rates estimated?
25. (*a*) Explain the meaning of maximum or optimum ownership level. (*b*) What factors determine this level? (*c*) Define new-owner demand and

state it symbolically. (*d*) Rewrite the basic equation for consumer durable goods demand, expressing these ideas.

26. Refer back to footnote 19, p. 177. What is the percentage of variation in private passenger car registrations accounted for by the independent factors in the equation? What is this measure called?
27. Compare the Commerce Department study of automobile demand with the Roos-von Szeliski study. (*a*) What economic factors, if any, do they have in common? (*b*) Were the approaches in both studies basically the same from an economic standpoint? (*c*) Compare the final results from a forecasting standpoint.
28. (*a*) When is the market for a durable good "saturated"? Explain fully. (*b*) What guides can you suggest to forecasting the sale of a consumer durable good, such as a gas range or washing machine, assuming the market for the good is already saturated?
29. (*a*) Explain briefly the essential variables to be considered in forecasting the sale of consumer durable goods. (*b*) The market-share approach is usually employed in arriving at a company forecast from an industry forecast. Explain, stating the assumption underlying this method and its significance for short- and long-range forecasting.
30. What essential variables should be considered in forecasting the over-all demand for capital goods, i.e., investment expenditures?
31. Cite some examples of pressure indices that may be used as an indication of management's desire to incur defensive investment in cost-saving capital equipment. What do we mean by "defensive investment," i.e., defensive against what?
32. What fundamental economic assumption underlies the attempt to derive formulas that will predict capital expenditures?
33. (*a*) What basic difference in approach do you see between the two steel studies on the one hand, and the portland cement study on the other? (*b*) How might an end-use index have been employed in the steel studies? How would the index be used for forecasting?

Chapter 6 PRODUCTION MANAGEMENT

After the foregoing study of demand forecasting and measurement, we turn our attention to the economic problems that confront managers in organizing and planning the firm's production and inventory. This represents a basic area of uncertainty to be minimized if management is to make correct decisions with respect to the employment of resources and the scheduling of output. An understanding of fundamental production relationships also provides a basis for the study of costs, which is the concern of the next chapter. For input-output relations in the company, or production functions as they are called by economists, involve the physical (technical and technological) conditions under which production takes place and hence are more basic than cost functions. Once the physical relationships between productive services and output are known, cost functions can be derived from production functions when the market prices of the productive services are given.

This chapter deals with three broad categories of problems. First, it outlines the basic physical relationships that exist between the inputs of resources or factors of production in a company and the corresponding output that results. This is treated in two sections, the first involving simple relations and the second dealing with multiple relations. Second, the economic and management policy aspects of a firm producing multiple products are considered. The problem here is to develop a framework for intelligent decision making with respect to a company's product line. Finally, a note on linear programming is presented in order to provide the reader with some insight concerning the implications of this science for problems of production management.

PRODUCTION FUNCTIONS: SIMPLE RELATIONS

Taking as synonomous the terms "resources" and "factors of production" to represent the inputs (men, machines, materials, etc.) required in production, the procedure followed below is to set forth the basic relationships between resources and products in a production process. By "production process" is meant the *transformation* of inputs into output. Such transformations of factors into product may occur within a single time period such as a year, or they may occur over several time periods, or they may never occur completely. The transformation (production) period thus varies between resources and thereby complicates the problems confronting the decision maker. In a static economy where the future is known with certainty, production would take place in a timeless vein without errors in estimation. But in the real world where uncertainty prevails, the recognition of time (uncertainty) excludes the possibility of perfect knowledge, and resources (representing investments over the years) must be analyzed for their effect upon output in terms of both fixed and variable costs. Time, and hence uncertainty, are the real causes of complexities in decision making with respect to a company's resource use.

The economics of production management takes as its starting point the study of the entire group of possible factor combinations that could be used to produce a certain output, within a given state of technology. The heading under which this type of analysis goes is that of the *production function*. A production function is an expression of the dependent or functional relationship that exists between the inputs (factors) of a production process and the output (product) that results. Hence it is also sometimes called the "input-output" relation. Like a demand function, a production function can also be expressed in the form of a schedule or a graph as shown subsequently, or algebraically by an equation such as $Y = f(X)$.

Realistically, the output of a product can never be ascribed to a single factor of production but is rather the result of combining several factors. A more accurate expression of the production function, therefore, would be $Y = f(X_1, X_2, X_3 \ldots X_n)$, where Y refers to the specific output as a function of the various input factors specified and unspecified. The only real requirement is that each of the letters represent a specific homogeneous class of factors. In this section we shall confine our attention to simple relations only; multiple relations will be considered in the following section.

The most elementary form of production analysis and the one which provides the basis for more complex consideration in production management is the single factor–product relationship. It is concerned with the transformation of a single input into a single output and hence for estimational purposes may be expressed conceptually by writing it in the form $Y = f(X)$. However, since the product Y will be the result of combining the input factor X (e.g., labor) with other factors (such as capital, land, management, etc.), the functional relationship may more appropriately be written $Y = f(X_1 \mid X_2, X_3 \ldots X_n)$. The vertical bar indicates that the input factors to the right are regarded as fixed in the production process under analysis, the factor to the left being varied. The fundamental problem in the study of the production function is to discover the probable nature of this input-output relationship. This is discussed in the literature of economic theory under several synonymous headings such as the "Law of Variable Proportions," the "Law of Diminishing Returns," or simply the "Laws of Return." Regardless of the name, it represents an explanation of one of the most widely held and best developed set of principles in economic science. The nature and ramifications of these laws are outlined briefly in the paragraphs below, based on a typical production function.

Auto Laundry Study

Figure 6–1, derived from Table 6–1, illustrates the results of a production function study for a small Detroit auto laundry.[1] The regression equation was a polynomial of the second degree fitted by the method of least squares; it took the form

$$Y = -0.8 + 4.5X - .3X^2$$

where Y represents total output in cars washed per hour and X is number of men. The study was based on 22 observations over a one-month period during which time the number of workers varied from a minimum of 3 to a maximum of 10. Given the equation, the total product data in column 2 of the table can be found simply by substituting values of X from 1 to 10 in the equation and computing the corresponding values of Y. The remaining columns can then be derived directly from this information and the three curves plotted as in the chart. Since no figures were available for less than three men, the curves were extrapolated as shown by the dashed portions and by the data in parentheses in Table 6–2. Incidentally, it should also be noted that the marginal product values in column 4 of the table are written midway between the X values in the table and plotted midway between the X values in the chart, since they represent the *change* in total output divided by the *change* in variable input ($\Delta Y/\Delta X$). A scale break (shown on the vertical axis) is also necessary to allow for greater readability of the average and marginal product curves.

Table 6–1 and Figure 6–1, especially the latter, reveal all the characteristics of the production function as they are typically expressed in standard works on economic theory. Hence, the important features of the chart, integrating both theory and measurement, may be summarized briefly as follows.[2]

---

[1] The study was financed by a firm interested in marketing a new type of mechanical washing system. A select group of auto laundries was chosen as a basis for the analysis.

[2] The bibliographical note at the end of the chapter provides references to several standard texts to which the reader can refer for more detailed treatment if his knowledge of production theory is rather limited. Therefore, only the bare outlines of the theory will be sketched here and in the following section.

First, the chart as a whole reveals the operation of the *Law of Variable Proportions* or the *Law of Diminishing Returns.* It shows that in a given state of technology, the addition of a variable factor of production, keeping other productive services constant, will yield increasing returns per unit of variable factor added, until a point is reached beyond which further additions of the variable factor yield diminishing returns per unit of variable factor added. This is the nature of the law as it is usually expressed in economics textbooks. It encompasses virtually all types of production functions from agriculture and automobile production through retailing and textile operations to the manufacture of zinc and zippers. It is thus a law of enormous significance as well as generality, as will become clearer from the discussion below.

TABLE 6-1

PRODUCTION FUNCTION FOR AN AUTO LAUNDRY
Regression Equation $Y = -.80 + 4.5X - .3X^2$

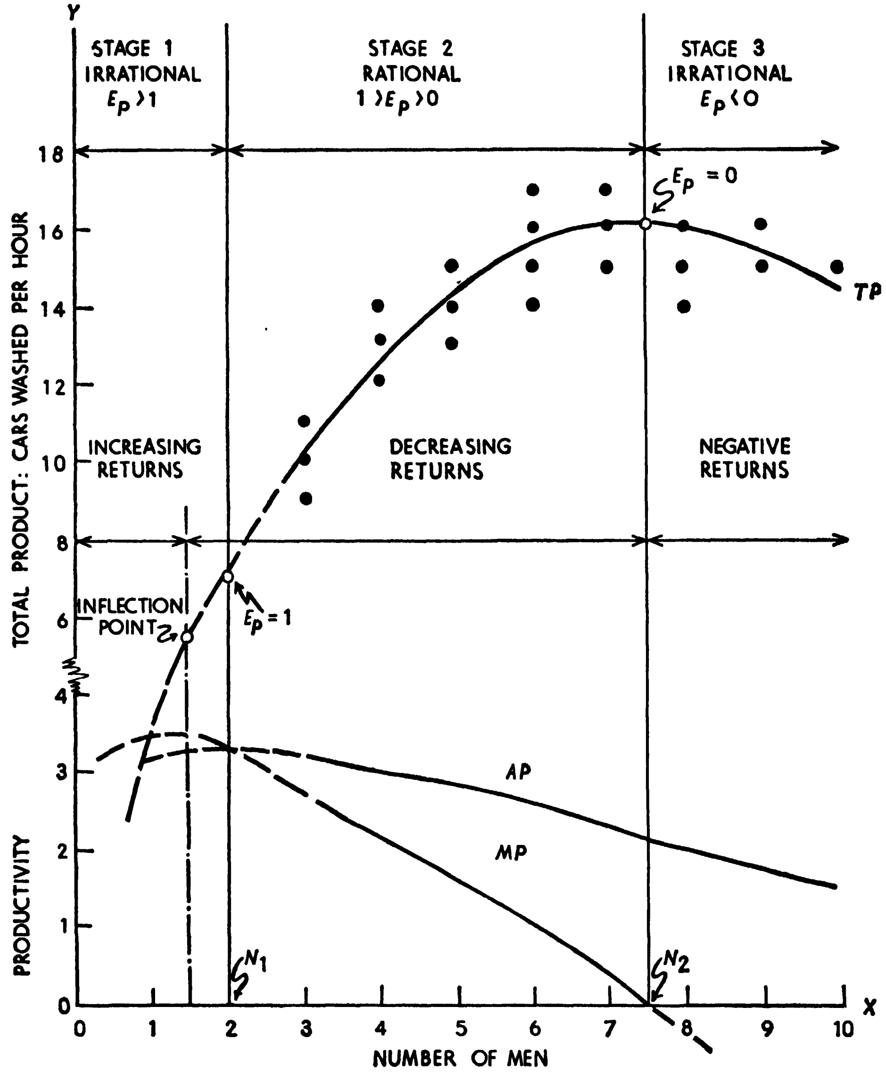
(1) Variable Input (No. of Men) X	(2) Total Product (Cars Washed per Hour) Y	(3) Average Product AP $AP = Y/X$	(4) Marginal Product MP $MP = \Delta Y/\Delta X$
(1)	(3.4)	(3.4)	
			(3.6)
(2)	(7.0)	(3.5)	
			3.0
3	10.0	3.3	
			2.4
4	12.4	3.1	
			1.8
5	14.2	2.8	
			1.2
6	15.4	2.6	
			0.6
7	16.0	2.3	
			0.0
8	16.0	2.0	
			−0.6
9	15.4	1.7	
			−1.2
10	14.2	1.4	

Source: See footnote 1

Second, the chart reveals what may be called the *total-marginal relationship.* The marginal productivity curve expresses the change in total product resulting from a unit change in input. Since total product is plotted on the Y axis and input on the X axis, marginal productivity is $MP = \Delta Y/\Delta X$. As long as this ratio $\Delta Y/\Delta X$ is increasing, i.e., the MP curve is rising, the total product curve is increasing at an increasing rate and is *convex* to the X axis. The point at which the TP curve changes

FIGURE 6-1

Physical Production Function for an Auto Laundry



Source: Table 6-1.

its curvature is the point of inflection and corresponds vertically with the peak of the *MP* curve as shown by the broken line in the diagram. In the Law of Diminishing Returns stated above, it is the peak of the marginal product curve that is referred to as the point of diminishing (marginal) returns—the point prior to which there are increasing returns to the variable factor and beyond which there are decreasing returns. (The peak of the average product curve represents the point of diminishing average returns.) When the total product curve reaches its maximum, at that point it is neither rising nor falling and hence its slope is zero. Since the ratio $\Delta Y/\Delta X$ also defines the slope of the total product curve, it follows that at that point the marginal product is zero. Beyond that point the total product is declining and hence must have a negative slope; the marginal product, therefore, is also negative, i.e., goes below the X axis. Increasing returns to the variable factor exist, therefore, when MP is positive and rising; decreasing returns occur when MP is positive and falling; and negative returns are realized when MP is negative and falling.[3]

Third, the chart reveals what may be called the *average-marginal relationship.* This is such that as long as the marginal product exceeds the average product, the average productivity of the variable factor increases; when the marginal product is less than the average product, the latter decreases; and when the average product is constant or a maximum, the marginal product is equal to it. A simple example illustrates this point. If to a class of students there is added a student whose age is above the average age of the class, the average age is increased; if his age is below the average age, the average decreases; if his age is equal to the average, the average remains the same. It should be noted from the diagram that even when the marginal productivity of the input turns down from its maximum point, the average productivity of the factor is still rising as long as its marginal productivity is greater than the average.

Fourth, economists customarily divide a production function of the type shown into what is known as the *three stages of production,* as illustrated in the chart. Stage 1 extends from zero input of the variable factor to where the average productivity of that factor is a maximum; stage 2 extends from the end of stage 1 to where the marginal product of the variable factor is zero (or to where total product is a maximum); stage 3 occurs where marginal product is negative (or total product is falling). Stages 1 and 3 are defined as *irrational* in that management, if it is to maximize profits, will never knowingly apply the variable to the fixed factors in any combination that will yield a total product falling in either of these two stages. The explanation is that stages 1 and 3 are completely symmetrical and hence the reasoning is as follows. In stage 1 the fixed factors are excessive relative to the variable factor and output

---

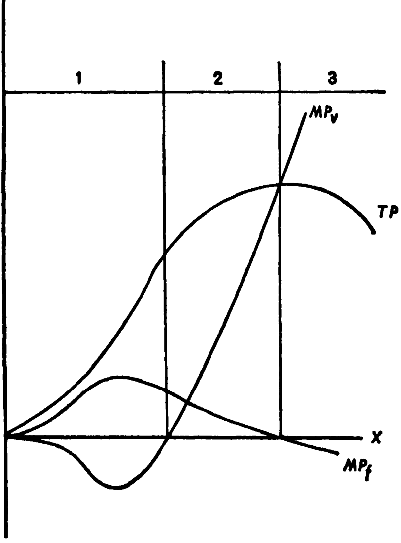
[3] It should be evident by now, at least to the reader familiar with elementary calculus, that the equation for any marginal curve can be derived by taking the derivative of the equation for the total curve. Thus, marginal profit is the derivative of total profit, marginal revenue is the derivative of total revenue, marginal product is the derivative of total product, and, as will be seen in the next chapter, marginal cost is the derivative of total cost. In general, for any marginal value M,

$$M = \frac{dY}{dX} = \underset{\Delta X \to 0}{L} \frac{\Delta Y}{\Delta X}$$

can always be increased by increasing the variable relative to the fixed (or by reducing the "fixed" relative to the "variable"). In a large department store understaffed with clerks, for example, sales can be increased by employing more clerks (the variable factor) relative to counters, floor space, etc. (the fixed factors), or by closing off sections of the store relative to the number of clerks. In stage 3 the variable factor is excessive relative to the fixed, and total output can be increased by reducing the variable relative to the fixed (or increasing the "fixed" relative to the "variable"). In the case of the department store again, if it were so overstaffed with clerks that they hampered each other or perhaps even kept customers from getting into the store and hence sales were declining, sales could be increased by reducing the number of clerks (or by increasing the size of the store). Stage 2 is the only *rational* stage of production, i.e., the only area within which profits can be maximized. Accordingly, management will seek to operate in the second stage because then neither input is being used in such excessive quantity as to reduce total output. Hence, the decision maker will employ a quantity of variable factor somewhere between N_1 and N_2 to maximize the economic returns of the firm.[4]

Fifth and finally, the chart reveals the elasticity of productivity (E_p) which measures the percentage change in output resulting from a 1 per cent change in variable input, and hence helps to explain the three

FIGURE C



The fundamental concepts underlying the three production stages can be further developed from Figure A in which are plotted total product (TP), the marginal product of the variable factor (MP_v), and the deduced marginal product of the fixed factor (MP_f). The diagram illustrates the symmetry of the relations. In stage 1 the marginal product of the variable factor is positive while that of the fixed factor is negative; in stage 3 the reverse is true. Only in stage 2 are both marginal productivity curves positive. If the variable factor is available free, the manager will go to the end of the second stage; if the fixed service is free, he will stop at the beginning of the second stage. The former principle is indicative of agricultural practices where labor is abundant relative to land, as in parts of the Far East; the latter helps to explain the lavish use of land by the colonists in early American history.

---

[4] The *precise* amount of factor hire will depend upon its price and the price of the product. The ratio of the two is the economic choice indicator which, when equated to the marginal product ratio, determines the maximum profit position.

stages of production outlined above.[5] In stage 1 the E_p coefficient is greater than unity (written $E_p > 1$) which means that a 1 per cent change in variable input brings a more than 1 per cent change in output. In stage 2 the percentage change in output is less than proportional to the percentage change in input but greater than zero (written $1 > E_p > 0$). In stage 3 where total product is falling, the percentage change in output is negative with respect to any percentage increase in variable input.

Meat-Packing Study

An alternative approach that might have been taken in the above study would have been to use man-hours instead of men as the independent variable. It was not done for the reason that the firm maintained a minimum staff of three men from Monday to Thursday and added further men on weekends when the demand for washed cars was greater. It thus adjusted to uncertainty fluctuations in output by varying the day-to-day requirements in its labor force, this being possible because the workers employed were neither skilled nor organized. But for most firms, it is impossible to make such short-run changes in the size of the labor force, although adjustments can more easily be made to predicted seasonal swings in output. However, flexibility can be built into the production function by recognizing that the labor force can be varied not only by the number of workers per shift, but also by the number of hours worked per day or per week, and by the number of shifts. Production functions can thus be derived in which output is a function of the labor force, the latter expressed in terms of its variable component dimensions of men and time.

An interesting analysis of production functions along these lines was done by William Nicholls for the fresh-pork operations of a large midwestern meat-packing plant.[6] In one of his analyses he arrived at the equation (based on 52 weeks of a single-shift operation during 1938–39):

$$Y = -2.05 + 1.06X - 0.04X^2$$

in which Y represents weekly total live weight (in millions of hogs processed) and X is weekly total man-hours (in thousands). The coefficient of determination (i.e., the proportion of explained variation to total variation) was $R^2 = 0.92$, which means that 92 per cent of the variations in the dependent variable are accounted for by variations in the independ-

---

[5] The point elasticity formula is:

$$Ep = (\Delta Y/Y)/(\Delta X/X) = (X\Delta Y)/(Y\Delta X) \gtreqless 1,$$

and can be used to measure the elasticity at any point on the *TP* curve either graphically or from the data in Table 6–1.

[6] W. Nicholls, *Labor Productivity Functions in Meat Packing.*

ent variable. As an alternative approximation of the production function, he fitted a logarithmic regression which took the form

$$Y = 0.39X^{1.12}$$

in which $R^2 = 0.87$. From these equations the corresponding average and marginal productivity curves can also be estimated and the results plotted graphically as was done above for the auto laundry study. Expressing the data as a logarithmic regression has a further advantage in that the exponent reveals the elasticity of productivity directly. Thus, for the above equation, a 1 per cent increase in total man-hours will result in an increase in total output of 1.12 per cent. The elasticity, therefore, is only slightly greater than unity.

PRODUCTION FUNCTIONS: MULTIPLE RELATION

The previous discussion concerned the factor-product or simple relations type of production function where output was assumed to be functionally related to one variable input. We turn our attention now to a more general type of production function of the form $Y = f(X_1, X_2 \mid X_3, X_4 \ldots X_n)$ or which, for purposes of measurement, can be considered as $Y = f(X_1, X_2)$. In this type of analysis at least two factors are considered variable; the relationships can then be generalized to include any number of variable inputs. The essential problem under consideration is to determine the minimum combination of variable factors for producing a given output, or the largest output that can be produced from a given combination of variable factors. Hence analyses of this kind involve a multiple relations type of production function.

An important law which differs from that discussed in the factor-product analysis of the previous section is the "Law of Returns to Scale." That is, instead of varying only one input and noting the effect on output as was done previously, we can consider the possibility of varying all inputs and measuring the change in output. For example, suppose that all of the factors in a production process could be varied in the same proportion, say doubled or trebled. It seems that the consequent change in total output would also be altered in the same proportion, being doubled if all inputs were doubled and trebled if all inputs were trebled. A relation of this type, where the percentage change in output is exactly proportional to the percentage change in all inputs as a whole (so that $E_p = 1$) is known as constant returns to scale, and a production function that exhibits this characteristic is said to be *linear and homogeneous*. In reality, conditions are rarely if ever encountered in which a production function is characterized by constant returns to scale over the full range of inputs, despite the fact that it might seem very plausible for constant returns to scale to be the rule rather than the exception. Actually, a production function would almost always exhibit alternating stages of both increasing ($E_p > 1$) and decreasing ($E_p < 1$) returns to scale due to two categories of phenomena explained below.

1. *Indivisibility of Productive Services.* The first condition which tends to prevent the occurrence of constant returns to scale over the full range of inputs is the indivisibility of productive services. Rarely is it possible to increase all of the productive factors in the same proportion; as a consequence, some of the factors are always being underworked or overworked relative to others at most levels of output, and this results in alternations of increasing and decreasing returns. For example, doubling the rate of output of an assembly line may still require only one final inspector instead of two; one locomotive may have sufficient horsepower to pull forty freight cars as adequately as twenty; a salesman may be able to take on a full line of goods instead of a single item at no significant increase in costs; and to a bank, the expense of investigating and managing a loan does not increase in proportion to the size of the loan. These examples from the fields of production, marketing, finance, etc., serve to illustrate that the advantage of size may result in economies that yield increasing returns to scale. That of decreasing returns to scale is illustrated in the following passage from a classic work on the subject:

> There is a story of a man who thought of getting the economy of large scale production in plowing, and built a plow three times as long, three times as wide, and three times as deep as the ordinary plow and harnessed six horses to pull it, instead of two. To his surprise, the plow refused to budge, and to his greater surprise it finally took fifty horses to move the refractory machine. In this case, the resistance, which is the thing he did not want, increased faster than the surface area of the earth plowed, which was the thing he did want. Furthermore, when he increased his power to overcome this resistance, he multiplied the number of his power units instead of their size, which eliminated all chance of saving there, and since his units were horses, the fifty could not pull together as well as two.[7]

2. *Decision-Making Role of Management.* The second factor tending to upset the plausibility of constant returns to scale lies in the decision-making role of management. In its function as coordinator, management may be able to delegate authority, but ultimately decisions must emanate from a final center if there is to be uniformity in performance and policy. As the firm grows, increasingly heavy burdens are placed on management so that eventually this class as a factor of production is overworked relative to others, and "diminishing returns to management" set in. Thus it is the growing difficulty of coordination that eventually stops the growth of any firm. As pointed out in Chapter 1, the development of sequential decision making as a science may have the effect of (*a*) reducing the time necessary to make a given number of correct decisions, or (*b*) increasing the number of correct decisions that can be made in a given time period. However, this would only tend to prolong the realization of decreasing returns to scale rather than eliminate it. Further, even if sequential-decision science could eventually overcome the limitational factor due to management, there is still the indivisibility

---

[7] J. M. Clark, *Studies in the Economics of Overhead Costs*, p. 116.

consideration discussed previously that serves as a major factor preventing constant returns to scale.

Cobb-Douglas Function

One of the pioneering econometric studies of production functions was done in the late 1920's by Paul H. Douglas, now United States Senator from Illinois and formerly Professor of Economics at the University of Chicago. Together with C. W. Cobb, he laid the groundwork in a 1928 journal article by deriving a production function for American manufacturing as a whole. The resulting analysis has come to be known as the "Cobb-Douglas function." It is probably the best known of numerous empirical studies that have since been done, and has served as an analytical basis for many subsequent econometric studies both in production and cost research.[8] In view of its importance, the results of the study are summarized below along with their applications to production management from the standpoint of the firm.

Using annual data for the United States based on the period 1900–1922, the Cobb-Douglas production function for American manufacturing took the form

$$P = 1.01L^{.75}C^{.25}$$

in which P is the production index of manufacturing output, L is the labor index of the average number of wage earners in manufacturing, and C is the capital index of fixed capital in manufacturing. All three indices are on an 1899 = 100 base period. The function is linear in logarithms but not in the original data as shown, so that the elasticities are given directly by the exponents. Thus, an increase of labor by 1 per cent results in a three fourths of 1 per cent increase in product; an increase of capital by 1 per cent results in a one fourth of 1 per cent increase in product. (In both instances, of course, other factors are assumed to remain constant). Productivity is thus seen to be relatively inelastic with respect to *each* of these two independent variables. An important *assumption* concerning the Cobb-Douglas function, however, is that the sum of the elasticities is 1. As a result, the production function as described earlier in the discussion of returns to scale is said to be linear and homogeneous: if inputs are doubled, output is doubled, and if inputs are trebled, output is trebled, etc., so that the elasticity of productivity with respect to the two inputs *combined* is unity. Does this mean that decreasing returns to scale will never actually set in, that small and large firms are about equally profitable, and therefore that the laws of returns to scale described previously are not actually valid? Not necessarily. Intuition alone would lead us to believe that decreasing returns to scale must eventually be realized, although possibly over a wide range of inputs, or else firms could continue to grow without limit. The economic explanation for the existence of constant returns as exhibited by the Cobb-Douglas function can probably be given by the fact that not all productive services were included in the analysis. At least one factor of production—management—was necessarily excluded from the empirical relationship, and it is likely that the omission of this scarce factor resulted in the appearance of constant rather than decreasing returns to scale as would otherwise have been expected.[9]

---

[8] See the bibliographical notes at the end of this and the next chapter.

Meat-Packing Study

In the meat-packing study mentioned earlier, Nicholls also fitted a multiple regression equation to the company data based on 52 weeks of single-shift operations during 1938–39. His analysis assumed a three-dimensional production surface (i.e., two variable inputs). Of six alternative multiple-regression equations that were fitted, the most general one resulting in a significant reduction of the unexplained variance was

$$P = -15.3 + 4.73M - 9.79M^2 + 5.02H - 0.44H^2 \qquad (1)$$

in which

P = weekly total live weight (millions of pounds) of hogs processed
M = weekly number of men (in hundreds)
H = average work week (tens of hours) per man

and for which the proportion of explained variation (the coefficient of multiple determination) was $R^2 = .96$. As an alternative to choosing this equation, the Cobb-Douglas type of function (a function which is linear in the logarithms) was also fitted to the data by Nicholls. He did *not* assume, however, that the sum of the exponents is equal to 1. There are at least two advantages to this type of function, i.e., a function which is linear in the logarithms, one of which has already been mentioned: the exponents reveal the elasticities directly; also, it allows for decreasing returns to be evidenced with the least complicated function. A function which is linear in the original data, on the other hand, would not do this. As in his other analyses of the company's input-output data, Nicholls fitted the Cobb-Douglas type production function by the classical method of least squares. The equation took the form

$$\log P = -0.84 + 0.95 \log M + 1.88 \log H$$

for which $R^2 = .92$. This regression may also be written

$$P = 0.14M^{.95}H^{1.88}. \qquad (2)$$

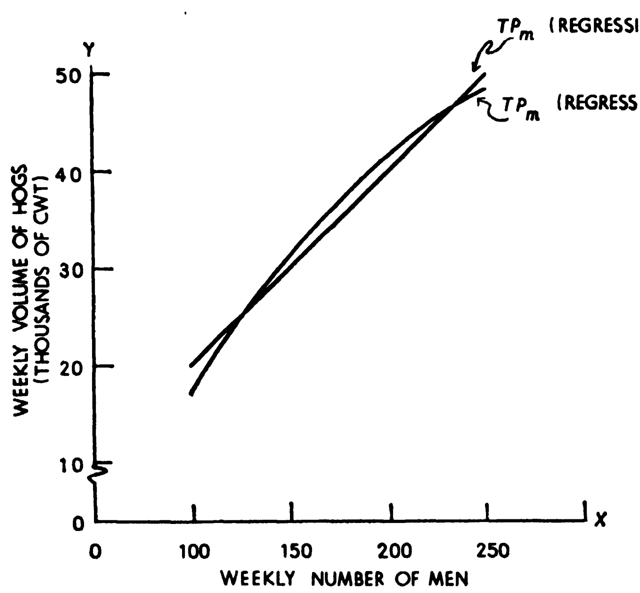
The latter equation shows that other things remaining constant, a 1 per cent change in the number of men, M, will bring a .95 per cent change in total product; also, a 1 per cent change in hours worked, H, is associated with almost a 2 per cent change in total output. From equations (1) and (2) the total, average, and marginal products can be derived, as was pointed out in the auto laundry study in the previous section, by substituting values for the independent variables and computing the corresponding values of the dependent variable. A production function in schedule form can thus be constructed and the data can be plotted graphically. Alternatively, the average product equations can be computed directly from the total product equations as illustrated in Table 6–2, and the desired results obtained in that manner. By substituting values for the independent variables in any of the equations, the corresponding dependent values can be easily computed. In Figure 6–2 the total productivity curves for number of men with work week constant at its mean ($H = 40.65$ hours), are graphed for values from $M = 100$ to $M = 250$. The corresponding average and marginal productivity curves can then be readily established, as described earlier for the auto laundry study.[10]

---

[9] The Cobb-Douglas function of the form $Y = aX^{b}_{1}X^{c}_{2}$ assumed constant elasticity of production, while a simple polynomial of the form $Y = a + bX - cX^2$ would allow for a negative marginal product. Either one may be satisfactory, however, depending on whether the range of observations includes an area of negative productivity. If it does not, the Cobb-Douglas function may be suitable; if it does, the simple polynomial may prove more satisfactory.

FIGURE 6–2

TOTAL PRODUCTIVITY OF MEN WITH HOURS CONSTANT, MEAT-PACKING FIRM



Source: Nicholls, *op. cit.*, p. 101.

---

[10] The marginal productivity equations, though not shown in Table 6–2, can be obtained directly by partial differentiation. Thus, with hours constant at their mean, the marginal productivity of men is:

(*a*) For regression (1): $\delta P/\delta M = 4.73 - 19.58M$

(*b*) For regression (2): $\delta P/\delta M = .133M^{-.05}H^{1.88}$

With men constant at their mean, the marginal productivity of hours is:

(*c*) For regression (1): $\delta P/\delta H = 5.02 - 0.88H$

(*d*) For regression (2): $\delta P/\delta H = .2632M^{0.95}H^{0.88}$

TABLE 6–2

Total and Average Productivity of Men and Hours, Respectively, Derived from Regressions (1) and (2)

Function	Regression (1)	Regression (2)
Total productivity of men and hours	$P = -15.3 + 4.73M - 0.79M^2 + 5.02H - 0.44H^2$	$P = 0.14M^{.95}H^{1.88}$
Total productivity of men*	$P = -2.19 + 4.73M - 0.79M^2$	$P = 2.03M^{.95}$
Total productivity of hours†	$P = -9.64 + 5.02H - 0.44H^2$	$P = 0.24H^{1.88}$
Average productivity of men*	$\frac{P}{M} = \frac{2.19}{M} + 4.73 - 0.79M$	$\frac{P}{M} = 2.03M^{-0.05}$
Average productivity of hours†	$\frac{P}{H} = \frac{-9.64}{H} + 5.02 - 0.44H$	$\frac{P}{H} = 0.24H^{.88}$

* Holding hours (H) constant at its mean ($\bar{H} = 4.07$).

† Holding men (M) constant at its mean ($\bar{M} = 1.72$).

Source Adapted from W. Nicholls, *Labor Productivity Functions in Meat Packing*, p. 100.

Derivation of Isoquants. From the regression equations it is possible, by fixing the amount of total product, to derive the relationship between men and hours for the given level of output. This is the factor-factor relationship discussed earlier, the plotted curve of which is an *isoquant* and represents the various combinations of input possible for yielding a given output. For example, suppose the firm forecasts an output of 34.6 thousand cwt. (3.46 million pounds) of hogs. What are the various combinations of factor inputs possible to yield this particular output? Taking equation (1) and substituting the value 34.6 for P, the equation becomes

$$34.6 = -15.3 + 4.73M - 0.79M^2 + 5.02H - 0.44H^2$$

or simply (combining 34.6 and 15.3)

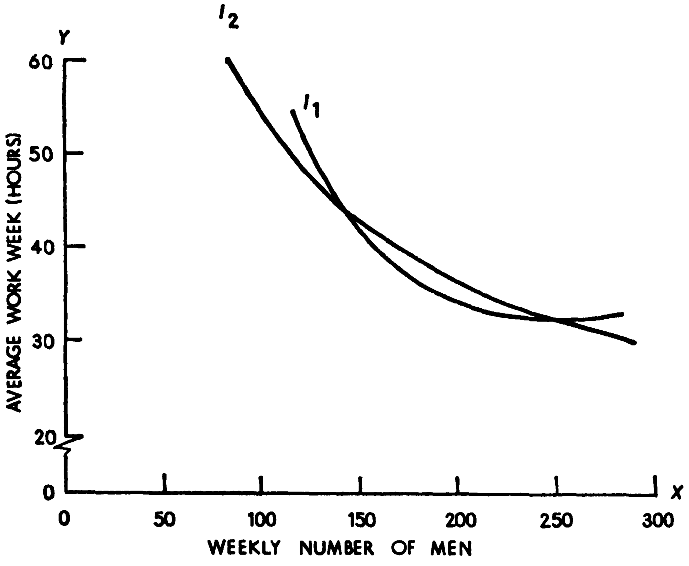
$$49.9 = 4.73M - 0.79M^2 + 5.02H - 0.44H^2.$$

Similarly, for equation (2)

$$34.6 = 0.14M^{.95}H^{1.88}.$$

FIGURE 6–3

DERIVED ISOQUANTS FOR MEAT-PACKING FIRM AT $P = 3.46$ MILLION POUNDS



Source: Nicholls, *op. cit.*, p. 105.

Either equation can then be solved for H, given various possible values for M, or can be solved for M, given various possible values of H. Figure 6–3 shows the respective isoquants for each equation derived in this manner, with I_1 representing the isoquant for regression (1) and I_2 the isoquant for regression (2), both at an output level of $P = 34.6$ thou-

TABLE 6–3

ISOQUANT SCHEDULES FOR SELECTED LEVELS OF OUTPUT (P), FOR A MEAT-PACKING FIRM
Derived from Regression· $P = 0.14M^{.95} H^{1.88}$

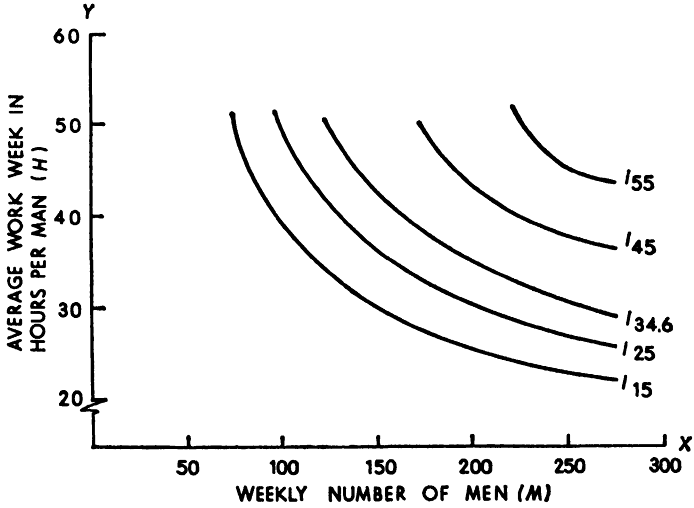
WEEKLY OUTPUT (P) IN THOUSANDS OF CWT.									
(1) $P = 15$		(2) $P = 25$		(3) $P = 34.6$		(4) $P = 45$		(5) $P = 55$	
Weekly No. of Men M	Hours per Man H	Weekly No. of Men M	Hours per Man H	Weekly No. of Men M	Hours per Man H	Weekly No. of Men M	Hours per Man H	Weekly No. of Men M	Hours per Man H
59	52.4	90	49.9	123	49.8	167	49.1	220	50.8
80	43.2	110	43.5	140	44.7	170	48.3	230	48.9
110	36.8	140	37.8	170	39.5	200	43.0	260	45.9
140	32.6	170	34.1	200	36.3	230	40.0	290	44.7
170	29.6	200	31.6	230	34.2	260	38.4		
200	27.5	230	29.8	260	32.9	290	37.6		
230	25.9	260	28.8	290	32.4				
260	25.0	290	28.3						
290	24.6								

Source Adapted from Nicholls, *op. cit*, p. 106.

sand cwt. Clearly, any number of isoquants could be derived at each level of output (P) by substituting values for M and computing H, or substituting values for H and computing M. Table 6–3 shows five different isoquant schedules computed in this manner, and the results graphed as in Figure 6–4. Each isoquant is indexed with its corresponding output

FIGURE 6–4

ISOQUANTS FOR SELECTED OUTPUT LEVELS, MEAT-PACKING FIRM



Source: Table 6–3.

level, and any point on any given isoquant shows the various combinations of men and hours capable of yielding a given output.

Decision Making

One of the most important uses of empirical production functions is to guide management in arriving at the most profitable level of factor hire. The goal, as stated earlier, is to attain *rational* production, which means: (1) obtaining a maximum output for a given collection of factors, or (2) obtaining a given output for a minimum aggregation of factors. The discussion of production functions in this and the previous section emphasized only the *physical* nature of the factor-factor and factor-product relationships, and at best was able to reveal only the range of optimum factor hire (e.g., stage 2 in a simple relations type of function) rather than the precise amount of factor hire for maximum profit. An extended treatment of the principles of production theory is beyond the scope of this book. Some brief comments may be made, however, so that the reader may gain perhaps an indication of the usefulness of empirical studies in this area.[11]

The physical production functions expressing the factor-product or factor-factor relationships can be converted into economic production functions as a guide for decision making once the relevant choice indicators are known. The choice indicators are the prices of the variables in the problem. In a factor-product (simple relations) function they are the price of the factor and the price of the product; in a factor-factor (multiple relations) function they are the prices of the respective factors. When the physical quantities are multiplied by their prices, the function is converted from a physical to a value function, i.e., the factor-product becomes a cost-revenue function. Expressed in this manner, management is then enabled to make decisions with respect to the level of factor hire by adjusting costs and revenues to yield maximum profits. In the factor-factor type of function, it is the factors that are variable and the problem is then to: (1) produce a given output for a minimum combination of factor costs, or (2) produce a maximum output for a given combination of factor costs. Regardless of the type of function, the goal is always to arrive at an optimum adjustment of factor units to product output, which is possible only when the nature of the production relations has been empirically established.[12]

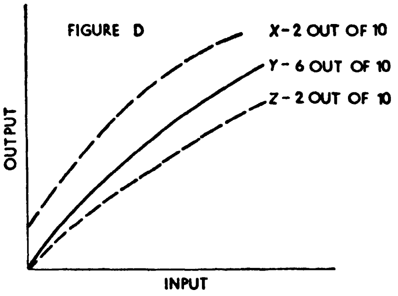
---

[11] More complete discussions are presented in the literature of economic theory. A good comprehensive treatment is available in T. Scitovsky, *Welfare and Competition*. Further sources are cited in the bibliographical note at the end of this chapter.

[12] Finally, it should further be pointed out that the effects of uncertainty may be introduced in terms of the discussion in Chapter 1, in the following manner. Instead of a single production function, a production planner may envision a distribution of production functions with a different probability value attached to each, based perhaps on past experiences. An illustration of three such curves is shown in Figure D along with their assigned probability values based, for example, on the past ten years of experiences.

If the efficiency (output-input) coefficients are randomly distributed, production function *Y* represents the mean and modal outcome since it was realized 6 out of 10 years, while *X* and *Z* represent the range or dispersion of outcomes for the functions higher and lower. We may then select a particular total product curve based on a mean or modal choice, and from this compute the corresponding average and marginal product curves. If the factor prices are known or can be predicted, we can also compute the relevant cost curves and thus arrive at a forecast of the entire cost structure based on mean or modal expectations of the future (see next chapter). Predictions over a period of ten years would thus approach the classification of a risk, while for any one year such predictions would be an uncertainty. If the distribution of efficiency coefficients is skewed rather than normal, we may of course prefer a mode to a mean expectation. Analogous concepts can also be framed for profit, demand, costs, and similar variables that form the chief areas of prediction and decision making in management planning.

FIGURE D



Conclusion

Obviously, the analysis of production functions is not "theoretical and impractical," for once such functions are empirically derived, they can aid management in exercising its coordinative functions of production scheduling and cost planning. Moreover, if input prices are known or can be forecast, and if the physical relationships between productive services and output are established in the form of a production function, the company's cost structure (cost functions) can be derived, and a foundation thereby laid for more effective profit planning. In the following chapter, dealing with cost functions, the significance of this will become more apparent. It might also be added that, in recent years, management scientists engaged in operations research have developed new approaches to the solution of certain production problems and, as a result, have opened up vast new areas for further research and analysis. Some of the implications of their findings are discussed briefly in the concluding section of this chapter, dealing with linear programming.

PRODUCT-LINE POLICY

The previous discussions in this chapter have centered around the problems that confront managers in achieving productive efficiency. Essentially, the analysis took the form of deriving quantitative measures of certain underlying physical input-output relations in a production process. We turn now to another aspect of production management: the problems of product diversification and specialization, i.e., the policy decisions that must be made by management in achieving an economic balance of its end products. In economic theory, problems of this type come under the heading of *multiple products;* businessmen, on the other hand, are more apt to use the term *product line.* In this and later chapters we use both terms interchangeably.

✓ Problems of multiple products fall into three broad categories for purposes of analysis: (1) product-line coverage or combination, which, as stated above, involves the establishment of policies for obtaining an economically balanced company output; (2) product-line pricing, which concerns the separate interrelationships between multiple product costs and multiple product demands; and (3) product-line improvement, which involves problems of a valuation nature over time. The first is discussed in this section, while the second and third are treated in later chapters on pricing and capital management, respectively.

Economic Bases of Multiple Products

In the final analysis, the reasons why management would be interested in expanding its product line are to increase profits and/or strengthen its market position with respect to competitors. The drive to expand profits by manipulating the product line is usually the result of excess capacity and thus places management on the offensive; keeping up with competitors, on the other hand, is more of a defensive tactic necessary for survival. Each of these is discussed further in the paragraphs below.

Excess Capacity. The presence of excess production capacity is perhaps the most important single factor prompting product-line diversification. If all productive services are not being fully utilized in an optimum manner, fixed costs are spread over fewer units and average total costs (i.e., unit costs) are thereby increased. The typical reaction, therefore, is to expand the product line, thereby reducing unit costs by obtaining a better utilization of capacity.

Excess capacity may occur for any of several reasons. It may be the result of an overly optimistic estimate of the market for the firm's product. In such cases, where growth is not anticipated to absorb this capacity in the reasonably near future, other products that can be readily adapted to the firm's plant and technical know-how might be added. As the firm matures and the question of new plant additions arises, the company's experience will enable it to determine what product or products to emphasize.

Excess capacity may be due to seasonal variations in demand, the latter being the result of weather and custom during the year. Classic examples are those of firms that diversify their product offerings in such businesses as coal and ice, shoes and rubbers, ice cream and dairy products, Christmas cards and birthday cards, etc. Companies faced with a seasonal demand for their products are certain to find it advisable to add existing products or develop new ones designed to mitigate the fluctuations in sales.

Overcapacity may be caused by cyclical fluctuations in sales. Products that are income elastic in demand, e.g., appliances, industrial equipment, etc., are particularly affected in recession and depression periods. Thus, Celanese Corporation of America, a synthetic textile manufacturer, has branched into chemicals and plastics to diversify out of the highly cyclical and too often unprofitable textile industry.

Excess capacity may result from secular shifts in markets, tastes, buying habits, etc., leaving the firm with underutilized facilities and know-how. The effect of the automobile on the bicycle and carriage industries is a common example of the changing pattern of market structures and product demands over very-long-run periods. Finally, the existence of excess or overcapacity may be a cause or result of vertical integration, a phenomenon that has become increasingly important in recent years and warrants some separate attention.

The causes of integration—the reasons why a firm may embrace a variety of products, markets, and functions—are attributable to a complex of historical, economic, and technical circumstances. The most obvious motive, however, is to get a strategic market advantage thereby reducing competitive uncertainties and enhancing profits. There is an economic motive to integrate whenever lower production costs will result. These reductions in costs may come about through a fuller utilization of plant capacity and other productive factors, or by creating new market opportunities. To the purchasing firm, the intermediate factors that it must buy from other companies in order to carry on its operations constitute a part of its costs. When the purchasing firm can supply its own resource needs more economically through integration than it can by securing these resources in the markets, profits will be increased by integrating. Similarly, in integrating forward, the additional (and average) costs at the new output level must not exceed the difference between the market price at the new and the old output level.[13] Frequently, in many instances a firm can produce two goods more cheaply together than separately: meat packers produce soap and other products from formerly wasted animal parts; grocery delivery trucks carry many products rather than a few; and salesmen can handle a full line instead of a few items at little extra cost (and, as a matter of fact, the products often complement one another thereby increasing sales). In chemicals, oils and electronics, long-term fundamental research has been particularly exploited by integration in an attempt to discover new and better product uses. These considerations, along with the importance of assuring sources of supply in order to maintain continuity of operations, have been major incentives accounting for the drive toward integration. The results, however, have often been dislocations in economies of scale along with technological and cyclical changes that serve to create an unbalanced product line from the company's profit-making standpoint.

---

[13] Space does not permit more than a few comments concerning the theory of integration. Some articles on the subject are cited in the bibliographical note at the end of the chapter.

Competition. Keeping up with (or ahead of) competitors is a second broad motive for product diversification. In industries where exact duplication is difficult or illegal due to secret processes, patent rights, etc., or where entry into the industry is relatively obstructed by economic or institutional obstacles, the need to produce similar (if imperfectly substitutable) products prevails. Industries characterized by monopolistic competition (many sellers, heterogeneous products) provide the most notable examples, though certain types of oligopoly structures are also illustrative. Profitable decisions in such circumstances require that the firm adhere to what has been called by some economists (e.g., Boulding) the "Principle of Minimum Differentiation": make the product as similar to competing products as possible without destroying the differences, thereby capturing part of the competitors' markets while at the same time instilling sufficient consumers' loyalty to minimize the shifts in buying due to minor price differences. The introduction by Armour and Swift of Treet and Prem, respectively, after the remarkable success of Hormel's Spam is an example; breakfast cereals, automobiles, and womens' clothing are further illustrations, the last two particularly as examples of product variation or differentiation in the short run due to style and fashion changes, and in the long run of product diversification as well.

Optimum Product Line

The previous discussion outlined the nature of the underlying pressures that prompt management to consider the production of multiple products. Granted that these pressures are sufficiently strong to cause diversification in production, what are the goals that may be adopted by the firm if it is to expand its output offerings? From a long-run standpoint, management may diversify its output in order to achieve an optimum product line that will maximize the profits from its resources. In the short run, on the other hand, it may decide to adopt as a "safeguard" against uncertainty a more immediate objective of income stability rather than profit maximization. In this case an optimum product line would be one which does not "place all of the eggs in one basket," in the hope that profits from the sale of certain products will offset losses incurred on others. In the long run, however, the goal of income stability merges with that of profit maximization and hence the concept of an optimum product line as one which fulfills this requirement remains essentially unaltered.

The economic problems facing the firm in making decisions as to what constitutes its optimum product line boils down to one of maximizing the return on its investment. The production of the goods represents an investment in resources (time, money, materials, etc.) which, over the life of the product line, gives rise to a series of expenses. Over the same period of the product-line's life, the firm will also incur receipts as the goods are sold. Thus production gives rise to a stream of costs over time and sales give rise to a stream of revenues over time. The difference between these two streams, the economic revenues less the economic costs, represents the stream of economic net profits or the return on investment, the present value of which is to be maximized over the life of the product line. The optimum product line is thus the combination of products that accomplishes this end. The firm is viewed as estimating a flow of revenues, costs, and profits over a time span and the decision is to select that combination of products which, out of all the alternative combinations, maximizes the present value of its income. When the problem is framed in this manner, product-line policy takes the form of making decisions in the present based on expectations of the future, and is thus a recognition of the uncertainty inherent in forward planning by management.

In view of the foregoing, the practical aspects of establishing the optimum product line resolves itself into two problems: (1) forecasting demand and costs for each new product to be added, and (2) arriving at a relevant concept of profit. On the revenue side, forecasting demand involves estimates of the product's price, advertising effectiveness, and other demand determinants discussed in previous chapters; on the cost side, the problems are largely those of predicting the labor, materials, and other expenses that will be incurred at given levels of output. (These are discussed in the next chapter.) For new products such forecasts are rarely more than pure speculation to begin with and are nothing but visions if they extend beyond a short-range (3- to 5-year) period, at least in the product's early stages of development. Establishing an appropriate profit concept means that, in principle, the incremental profit attributable to the addition of the new product should exceed the incremental returns that would be incurred by investing the resources in the next best alternative use; in practice, this means that the income expected over the life of the product less the outlay and investment (i.e., economic cost) must be greater than the income received from any other investment alternative, including considerations of discounting and compounding. Problems of this and a related nature, which involve investment decisions over time, form part of the subject known as "capital budgeting" and are treated in further detail in later chapters. Our purpose at this point has been to emphasize that the relevant profit concept is one that is expressed in a form which permits comparisons to be made with alternative uses of the same resources (labor, plant, materials, etc.) over time. The cost to the firm is thus measured by the sacrifice it incurs or the return it foregoes by not using these resources in their most profitable alternative.[14]

---

[14] Products are thus portrayed as competing for the firm's limited resources; hence, the market cost of new capital to the firm should be the true profit standard

Product-Line Expansion

If greater net profits are realized from additions to the product line, it may frequently be the result of at least two separate but related underlying factors: one of these is *product interdependence;* the other is *excess capacity.* Each of these represent a class of causes in many and varied forms that may be sufficient to prompt a product-line expansion.

Product Interdependence. The relationships that may exist between the products of a multiple product firm may be of a competing, complementary, or independent nature. Competing (substitute) products, from a demand standpoint, often serve as a precaution against uncertainty by reducing the probability of sales variations due to changing demands, tastes, etc., while recognizing that satisfactions rather than products are what buyers are purchasing. In other words, a company must always face the real possibility of product obsolescence either by itself or by competitors, and firms that offer a wide array of substitutes in their respective product lines (e.g., Lever Brothers, Proctor and Gamble, etc.) are frequently hedging against this type of uncertainty.[15] Complementary products, on the other hand, provide the firm with a profitable opportunity to fill the related needs of the buyer (e.g., shoe stores selling polish, hose, etc.). The hope on the part of management in expanding the product line is either that the existing products are well enough known to sell the new product, or that the new product will perhaps excite enough sales to increase the demand for the existing products. Finally, when products are independent, they are supplementary to one another and have no direct effect on the sale of other products by the firm. From a production standpoint, however, all products in the line, regardless of the relationships between them, compete for the firm's resources (including management) and may thereby raise important opportunity cost considerations in deciding on the optimum product line.

Excess Capacity. Product-line expansion to utilize available capacity may be motivated by a variety of considerations. Thus it may be the result of an effort to fill in seasonal dips in sales, as when a firm sells air conditioners and heaters; it may be the result of utilizing common advertising media and distribution channels, e.g., supermarkets; it may be the result of management's desire to provide a "full line" to customers by utilizing the excess capacity that exists in the company's brand name or reputation, as with Ford's introduction of the Edsel. Numerous other possibilities exist

---

as will become evident in the later chapter on capital budgeting. When resources are not limited, a profit rate based on historical average earnings for a past period may be quite practical. See also Chapter 4 on the discussion of profit standards and the meaning of "incremental profit."

[15] Buggy manufacturers, by way of contrast, failed to recognize that they were selling transportation rather than carriages, and were replaced by the automobile industry as a result.

and the reader can undoubtedly think of several. In any event, the considerations in expanding the product line involve, from the standpoint of excess capacity, questions of both productive and distributive efficiency.

Production aspects include: (1) the integration of existing facilities for the new product in the form of plant space, machinery, etc., with seasonal and/or cyclical variations in production; (2) the proportion of the firm's present resources to be allocated to the product and the proportion to be acquired from outside (e.g., should the company manufacture the product and farm it out for assembly?); and (3) the availability of materials in sufficient quantity to assure required output levels. To a large extent these factors are conditioned by the existence of common production facilities for the new product and the existing product line, thereby making possible a fuller utilization of excess capacity and thus the sharing of overhead costs.

Distribution aspects involve: (1) the place of the product in the company's regular distribution pattern (e.g., whether the product can be sold through the same or regular channels, which will be partly determined by the opinions of the product held by jobbers and wholesalers); (2) whether the present sales force can handle the new product without prohibitive increases in salesmen's costs and time, and without neglecting other products; and (3) the amount of advertising and promotion needed in the product's introductory and early growth stages. Integrated with both the production and distribution aspects are also certain financial considerations such as manufacturing costs, sales and advertising costs, capital requirements, inventory levels to be maintained, pricing methods, and profit planning. Some of these have already been discussed in early chapters; others are treated in more detail in this and subsequent chapters.

Product-Line Contraction

For the most part, the converse of the rules stated above with respect to product-line diversification also apply to product-line contraction. In principle, the optimum product line is the one which yields the greatest long-run rate of return for a given investment of resources, or yields a given long-run rate of return for a minimum investment of resources. Hence a product can be dropped if the same resources used to produce it could be used more profitably in a better alternative, provided that net returns on the company's total resource investment would be thereby increased. Usually, if a product is not evidencing profitable performance, management can consider the three alternatives of make, buy, or drop.

1. *Make.* If the firm continues to make the product, it may require an improvement in production and/or distribution efficiency as outlined above to yield adequate returns. If the commodity is a by-product, it may be sufficient to retain it as long as its contribution profit (revenue minus variable cost) is positive. Advertising, promotion, and other selling expenses may even have to be minimized for the product in order to raise

the contribution margin. A further cut in distributive costs might be realized if the product were manufactured by the firm and farmed out to others for final sale. The practices of Sears Roebuck, Montgomery Ward, and other mail-order houses are notable examples of companies that assume the marketing functions which manufacturers may not otherwise be equipped to handle.

2. *Buy.* A decision to buy the product rather than make it is justifiable if the supplying firm can provide the product in sufficient volume and at low enough costs to make it sufficiently profitable for resale by the buying company. For the buying firm this alternative has the following consequences: (*a*) it makes the firm more dependent on others, which may not be a disadvantage if the supply of the product can be assured; and (*b*) if the supplying firm is part of an oligopolistic industry, its pricing practices may be sufficiently erratic to complicate profit, cost, and sales planning by the buying company.

3. *Drop.* The decision to drop the product entirely is warranted if its long-run net profit is below what would be attained from an alternate product using the same resources. It is important to emphasize long-run and not short-run profits (or even losses). In the short run, only contribution profit is the relevant consideration, since earnings above variable expenses go to sharing the overhead and the earning of profit. In the short run, resources are a sunk cost and any spreading of fixed expenses over more products is justifiable. Long-run considerations, however, allow for greater resource mobility and hence all fixed costs become variable. (This is discussed further in the next chapter). Net profits resulting from various alternatives thus become the only relevant criterion as a basis for decision making.

OPERATIONS RESEARCH—LINEAR PROGRAMMING

Various approaches to the solution of problems in production management, particularly approaches employing mathematical procedures, have been developed in recent years by management scientists or operations researchers. The fundamental aim has been to introduce rigorous methods to the analyses of business problems and the process of management decision making under conditions of uncertainty. From an abstract standpoint, operations research (i.e., the application of mathematical methods to business problems), decision theory, and economic analysis (including econometrics) are really identical. The analytical procedure in all instances consists of four parts: (1) arraying the alternative possible goals to be sought, (2) defining the assumptions to be employed, (3) determining and balancing the net advantages and disadvantages in selecting the optimum goal, and (4) modifying the selection by recognizing the institutional factors both inside and outside the firm that might make certain choices "impractical" or otherwise unpalatable, so that the final choice will fit in properly with the firm's over-all objectives.

Where the approach is mathematical, a model is typically constructed which incorporates the empirical relations that are relevant, the assumptions to be employed, and even the objective to be sought. The solution may be adopted directly or modified in some way to fit better the firm's over-all objectives.

An extended treatment of operations research (OR) is beyond the scope of this book, but a few brief comments may be made at this time to indicate its role in dealing with certain problems in production management, particularly linear programming, optimum product lines, and inventory.

Linear Programming

When the production process in a firm can be broken down into a series of "straight-line" relationships, this facilitates studies of the budgeting type whereby the object is to predict the cost-returns effect of certain readjustments in resources. Econometricians and operations researchers (essentially the same) have given the name of "linear programming" to studies of this type. For example, in deciding on the most profitable products to produce, linear programming would apply if a doubling of all inputs (labor, material, etc.) will approximately double the output, as would occur if two factories with twice the labor force, etc., could produce twice the output of one factory and production is increased by building more, rather than larger, plants. Many production processes in actuality do involve linear relationships when the plant as a divisible producing unit is concerned. This statement, of course, does not apply to single resources which are indivisible for technological reasons: a railroad cannot lay one and one-half tracks; dyes must be used in complete sets; etc. Yet, if all the resources (equipment, facilities, etc.) which make up a plant as a whole are considered, they do frequently exhibit a series of straight-line steps or relationships between input and output if the plant is small enough so that decreasing returns to scale are not encountered. Each alternative technique or method for completing the step will yield a particular input-output ratio, and by comparing the ratios to the factor-commodity price ratio, the best procedure can then be adopted and applied to all technical units. This type of "process analysis" thus consists of the fitting together of a series of linear processes for the purpose of solving a general class of optimization problems in the field of production and related areas.[16]

In general, linear programming is applicable where a given percentage increase in all the independent variables is just sufficient to permit about an equal percentage increase in all the dependent variables, i.e., where it costs ten times as much to produce ten items than it does to produce one, measured in terms of production time, etc. If this assumption is unrealistic, or if the functions cannot be linearized by transforming the variables (e.g., using logarithms, powers, etc., as explained in previous chapters), then linear programming will not be applicable.[17] It may happen that economies can be effected by increasing factory size, in which case production increases faster than inputs. The graph of the relationship between total inputs and outputs will not then be a straight line (i.e., the functions may be quadratic in nature), and a more complicated mathematical analysis known as "nonlinear programming" may be applied. In practice, however, the procedure is often to employ linear programming because of its greater simplicity and the fact that it often yields sufficiently close approximations to the correct answer.

---

[16] Linear relationships are particularly notable in farming. For example, the preparation of a seedbed for corn is of this nature, in that each added acre plowed, unless the size of the farm is very large, requires an equal input and adds about an equal increment to output; and similarly, corn planting, corn cultivation, and corn harvesting often involve linear input-output relationships. (*Cf.* E. O. Heady, *Economics of Agricultural Production and Resource Use*, p. 83.) Manufacturing offers many illustrations as well.

Optimum Product Lines

Programming techniques have been successfully applied to the solution of a variety of problems involving the selection of optimum product lines. If a firm is faced with limited capacity, the real cost of making an item includes not only the manufacturing costs but in addition the loss of profits (opportunity costs) resulting from a decision not to release capacity for the production of other, more remunerative lines. Further, there is the problem of defining a "most profitable" item. One item may make optimum use of scarce machine time and thereby yield maximum profit per machine-hour, while another product may make optimum use of limited warehouse space. A decision to produce only the former product would result in full utilization of warehouse facilities while machine time was still underutilized; production of the latter product, on the other hand, if it were not bulky, might result in nearly full utilization of machine time and leave warehousing facilities substantially underused. There is thus an interaction of variables subject to restraining conditions (e.g., capacity limitations, minimum amounts required for each product, etc.), and the methods of solution fall into a general class of optimization problems that are feasibly handled by programming procedures.

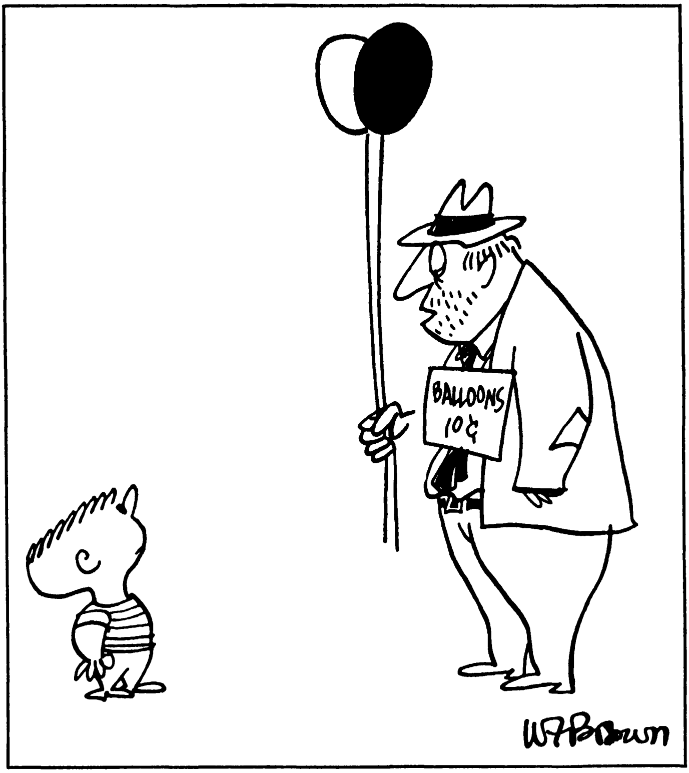
Inventory

The principal causes for the existence of inventories are: (1) expected changes in demand or cost functions, (2) discontinuities in ordering, production, and sales rates, and (3) demand uncertainty. Economic theory has been concerned mainly with the first of these factors, leaving the remaining two virtually untouched. Yet, an analysis of these last two would fill an important gap between economic theory and business practice.[18]

---

[17] A curvilinear relationship may be approximated, however, by isolating many linear segments on a curve, i.e., the objective function is piecewise linear. The assumptions of linear programming are those of linear relationships or discontinuous functions.

The problem of determining optimum inventory levels is a matter of balancing advantages and disadvantages. Too low an inventory results in higher shipping costs and in increased likelihood of lost orders when customers must turn elsewhere; too high an inventory results in high storage costs, insurance costs, and tax and interest payments on money capital tied up in inventory. Standard OR methods have been developed for arriving at optimum inventory levels (within the limits permitted by the data) by first translating the various cost relationships into mathematical form and then employing differential calculus to determine the inventory level that will maximize profits. In addition, programming computations are employed to yield intermediate objectives such as evening out produc-



"High interest rates coupled with a slight decline in gross national product forced me to trim inventories. That's why there ain't no green one, sonny."

---

[18] Recent studies in this direction have been made by W. Eiteman, *Price Determination, Business Practice versus Economic Theory*, pp. 67–68; and K. Boulding, *A Reconstruction of Economics*, pp. 111 ff., but both leave much to be desired. For a brief summary and evaluation, see T. Whitin, *Theory of Inventory Management*, chap. 4.

tion fluctuations, meeting unexpected customer demands, and expediting the anticipated flow of products from factory to market in a way that will minimize costs.

Conclusion

In addition to the development of OR methods for handling problems of the nature outlined above, there are further areas in business administration where similar analytical procedures have been employed successfully. These include among other things the problem of determining optimum plant and warehouse locations which, to a large extent, involves the minimization of costs in transporting raw materials and finished products, with considerations of labor supply and the speed of servicing markets taken into account. Evidently, scientific methods for handling many kinds of management problems do exist, and econometrics, which has occupied most of our attention, is only a part of this broad science of operations research.

BIBLIOGRAPHICAL NOTE

General and special studies of the production function are quite numerous, of which the following are suggestive: S. Carlson, *A Study in the Pure Theory of Production;* J. M. Cassels, "The Law of Variable Proportions," in *Explorations in Economics in Honor of F. W. Taussig;* O. L. Williams, "Suggestions for Constructing a Model of a Production Function," *Review of Economic Studies* (1933–34); and discussions of economies of scale by A. N. McLeod, F. H. Hahn, and E. Chamberlin in the *Quarterly Journal of Economics* (1949). Well-known empirical studies include P. H. Douglas and Cobb, "A Theory of Production," *American Economic Review* (1928), and a number of subsequent studies by Douglas with others applied to Canada, Australia, and the United States appearing in various issues of the same journal as well as the *Quarterly Journal of Economics* and the *Journal of Political Economy* until 1941; at the microeconomic or firm level, W. Nicholls, *Labor Productivity Functions in Meat Packing* is well worth consulting. A classic article stating the famous controversy on increasing returns is P. Sraffa, "The Laws of Return Under Competitive Conditions," *Economic Journal* (December, 1926), and reprinted with some introductory comments in P. C. Newman, A. D. Gayer, and M. H. Spencer, *Source Readings in Economic Thought*, Part 14. Finally, on the economics of excess capacity and integration, two useful sources are N. Kaldor, "Market Inperfections and Excess Capacity," *Economica* (1935); and G. Stigler, "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy* (June, 1951).

Readers desiring more comprehensive and less technical treatments will find concise theoretical discussions of the production function in most textbooks on intermediate economic theory. Perhaps the simplest is M. Bober, *Intermediate Price and Income Theory*, while more advanced treatments appear in G. Stigler, *The Theory of Price*, rev. ed., and S. Weintraub, *Price Theory*. Applications to agriculture, which, however, find their counterparts in industry as well, are found in various portions of Heady's excellent work mentioned previously. On product diversification, see Colberg *et al.*, chap. 8; Dean, chap. 3; H. Hansen, *Marketing: Text, Cases, and Readings*, chap. 2; and D. Phelps, *Sales Management*, chap. 2. Some indications of the scope of operations research are presented in: Alderson and Sessions, *Cost and Profit Outlook*, May, 1956 and March, 1957; in American Management Association, *Operations Research, A Basic Approach* (Special Report No. 13); and in a recent report on the subject by the National Industrial Conference Board, *Operations Research* (Studies in Business Policy No. 82).

QUESTIONS

1. What is meant by the term "production function"?
2. The underlying concept of a production function is the law of diminishing returns, more technically termed the law of variable proportions. Why? Explain.
3. In one of the auto laundry studies mentioned in footnote 1, p. 204, the estimating equation turned out to be $Y = -.93 + 3.52X - .27X$, where Y represents cars washed per hour and X is the number of men. Construct a table as in Table 6–1, p. 205, for values of X from $X = 1$ to $X = 9$. Plot your results on graph paper and label the curves, stages, and other relevant characteristics as in Figure 6–1, p. 206. The equation was derived from the following observations:

X = no. of men	1	2	3	4	5	6	7	8	9
Y = cars washed per hour	2	6	7	8	10	11	11	10	9

Note that these points should be plotted as in Figure 6–1. The above equation represents the line of best fit for these points. Values of Y may be rounded to the nearest tenth in constructing the table.

4. In your answer to question 3, account fully for the shape of the curves and the items labelled on your chart, by explaining the following: (*a*) the law of variable proportions; (*b*) the total-marginal relationship; (*c*) the average-marginal relationship; (*d*) the three stages of production and their reasons for existing; and (*e*) the elasticity of productivity.
5. What basic difference existed between the auto laundry study and the meat-packing study? Of what significance is this?
6. What is the effect of a technological improvement on a company's production function?
7. It has been said that were it not for the law of diminishing returns, it would be possible to grow all of the world's food in a flowerpot. Explain. (*Cf.* K. E. Boulding, *Economic Analysis*, 3rd ed., p. 603.)
8. What is meant by the law of returns to scale? Is this law basically analogous to the law of diminishing returns?

9 (*a*) What conditions usually operate to prevent the realization of constant returns to scale? (*b*) Typically, at what levels of operations, i.e., small or

large, would you expect increasing returns to scale? Decreasing returns to scale? Why?

10. (*a*) What is an isoquant? (*b*) In general, what is the value of deriving empirical production functions?
11. (*a*) Why do firms expand their product line? (*b*) Of what significance is excess capacity in product-line policy? (*c*) What are the common sources of excess capacity?
12. (*a*) What is meant by integration? (*b*) Why do firms integrate? (*c*) In what way may costs be reduced by integration?
13. (*a*) Explain the meaning of "optimum product line." (*b*) What forecasting problems are involved in arriving at an optimum product line?
14. Explain the appropriate profit concept to employ in decisions on adding new products.
15. (*a*) What types of production and distribution considerations are encountered in expanding a product line? (*b*) In product line contraction?
16. In your own words, formulate a definition of operations research from a business standpoint.
17. How does econometrics relate, if at all, to operations research?
18. In what fields, from a business economist's standpoint, has operations research made the greatest headway?

*

Chapter 7

COST ANALYSIS

In his classic work on overhead costs, J. M. Clark makes the statement, "a graduate class in economic theory would be a success if the students gained from it a real understanding of the meaning of cost in all its many aspects."[1] The area of cost is certainly one of the most complex of applied economics and at the same time is one which has occupied the attention of economists for many years. To bring out the full accounting and economic implications, not to mention the engineering aspects, would require many chapters if not an entire book. The following sections, therefore, will merely sketch the more essential characteristics of cost analysis with the main emphasis on cost theory and empirical measurement. As in previous chapters, the chief concern is with the formulation of concepts that can be of use to managers in coordinating the firm's activities and facilitating forward planning.

NATURE AND THEORY OF COST

The general idea of cost covers a wide variety of meanings, but there is one meaning that is common to all types of costs and is summed up in the single word "sacrifice." The nature of the sacrifice may be tangible or intangible, objective or subjective, and for this reason a chief difficulty for decision purposes is to represent costs by appropriate numbers that can be readily manipulated in the accounts. It is common, therefore, to avoid such concepts as psychic costs or sacrifices in the form of mental dissatisfactions, social costs such as the smoke nuisance of a factory, and sometimes real costs, e.g., sacrifices in purchasing power. Actually, most of the controversy over the existence of various kinds of costs evaporates once it is realized that there are different kinds of problems for which cost information is needed, and that the particular information required varies from one problem to another. The fact that accountants, economists, and engineers are each concerned with the study of costs for different purposes explains why there is a large variety of ideas about costs, many of which are adapted to different purposes. The following classification of some common cost concepts will help fix certain basic ideas and relations.

---

[1] J. M. Clark, *Studies in the Economics of Overhead Costs*, preface, p. ix.

Classes of Costs

A classification of important cost concepts should dispel immediately the notion that conventional accounting practice provides the firm with all its necessary cost information, and should drive home the fact that cost concepts differ depending on managerial uses and viewpoints. In practical work, the historical costs provided by accounting are often sufficient to fulfill certain legal and financial requirements, but for economic decision making where the concern is with predicting costs under alternative courses of action, conventional accounting usually leaves much to be desired. As will be seen below, the most useful estimates are frequently those that are derived by combinations and adjustments in the data, evidencing the fact that in the well-managed firm the accounts are a source of basic information rather than an end in themselves.

1. ***Absolute Costs and Alternative Costs.*** One of the most fundamental distinctions between two general classes of ideas of costs is that between *absolute* or *outlay costs* and *alternative* or *opportunity costs.* Absolute costs involve an outlay of funds or, in fact, all reductions in assets such as wages paid, materials expense, rents, interest charges, and so on. Alternative cost, on the other hand, concerns the cost of foregone opportunities, or in other words a comparison between the policy that was chosen and the policy that was rejected. For example, the cost of lending or using capital is the interest that it can earn in the next best use of equal risk. The alternative uses of capital measure the marginal cost of capital to a given borrowing or lending firm. If capital funds can earn 5 per cent in their most productive employment, then that is their cost to the firm employing the funds. Similarly, assuming full capacity operations, the cost of a product in the product line is not merely the outlay on resources, but also the profit that would have resulted from the best alternative product produced with the same facilities. Evidently, the basis of choice or decision where alternative costs are involved hinges on a comparison between what the firm is doing and what it could be doing, and it is the difference between those alternatives that constitutes the critical cost consideration.

A subdivision of opportunity or alternative cost is *imputed costs.* These never show up in the accounting records but are nevertheless important for certain types of decisions. Interest (never paid or received) on idle land, depreciation on fully depreciated property still in use, interest on equity caital, and rent on company-owned facilities are examples of imputed costs. To illustrate, in evaluating the relative profitability of two warehouses owned by a company in order to decide whether to continue, discontinue, or lease them requires supplementary calculations of rent and interest on investment. In deciding how much to impute to each, the answer rests on the uses to which the space released could be put and the relative profitability of each use. Although precise calculations are rarely possible, the concept nevertheless provides for a correct way of thinking about such problems and a basis for establishing at least rough estimates for better decisions.

2. ***Direct and Indirect Costs.*** *Direct costs* are costs that are readily identified and visibly traceable to a particular product, class of products, operation, process, or plant. The concept may also be extended beyond the sphere of manufacturing costs; thus, overhead may be direct as to departments, and manufacturing costs are frequently direct as to product lines, sales territories, customer classes, and the like.

Indirect costs are costs that are not readily identified nor visibly traceable to specific goods, services, operations, etc., but are nevertheless charged to the product in standard accounting practice. The importance of the distinction between direct and indirect cost from an economic standpoint is that some indirect costs, even though not traceable to product, may nevertheless bear a functional relation to production and vary with output in some definite way. Examples of such costs are electric power, heat, light, and depreciation based on output.

Practically synonymous with indirect costs are *common costs.* Common costs are costs that are incurred for the general operations of the business and yield benefits to all products, e.g., the president's salary. When the outputs involved are related to each other, common costs are also known as joint costs. Thus, the cost of crude petroleum is common to gasoline, kerosene, etc.; the cost of raising cattle is common to the yield of beef and hides.

The significance of cost traceability becomes considerable when management must make decisions in the areas of product-line policy (Chapter 6) and pricing policy (Chapter 8). Most industrial firms produce multiple products for which there are at least some common costs, but for which there may be substantial differences in production and marketing processes. Cost tracing all the way back to the individual product is not necessary and is not the basis for intelligent pricing policies in such firms. Instead it is sufficient for management to know the separate cost of classes of output and to price on the basis of typical conditions rather than accidental variations resulting from irregularities and imperfections in various resources and markets.

3. ***Fixed Cost and Variable Cost.*** Economists generally distinguish between two major categories of cost as fixed cost and variable cost. *Fixed costs,* or "constant" costs as they are sometimes called, are those costs that do not vary with (are not a function of) output. They are costs that require a fixed outlay of funds each period such as rent, property taxes and similar "franchise" payments, interest on bonds, and depreciation when measured as a function of time (without any relation to output). It should be emphasized that the term "fixed" refers to those costs that are

fixed *in total* with respect to volume, for they may still be a function of capacity and hence vary with plant size. In other words, fixed costs are not fixed in the sense that they do not vary; they may vary and frequently do, but from causes that are independent of volume. It follows of course that since fixed costs are constant in total, they will vary per unit with the rate of output, continuously decreasing as output increases within the production period.

A term synonomious with fixed cost, at least to the economist, is *overhead cost.* To the cost accountant, however, the meaning of this term is virtually the same as indirect cost. Overhead, in accounting literature, usually is composed of some fixed costs and some costs that are variable in nature. The distinction is unfortunate and can lead to misinterpretations in technical discussions if care is not taken in defining terms.

Variable costs are those costs that are a function of output in the production period. Unlike fixed costs the resource services of which are given off in a constant flow irrespective of the output quantity, variable costs emanate from the stock services that are transformed or "used up" as output is produced. Variable costs vary directly, sometimes proportionately, with output. Over certain ranges of production they may vary less or more than proportionately with output depending on the utilization of fixed facilities and resources (discussed further below). The sum of these two categories of cost, fixed and variable, at any given output level, yields the total cost at that output level, i.e., $FC + VC = TC$. When derived for successive levels of output, the resulting TC series represents a functional relationship between total cost and output, as will be seen later in this chapter in the discussion of theoretical and empirical cost functions of various kinds. Examples of variable cost include materials utilized, power, direct labor, factory supplies, salesmen's commissions, and depreciation on a production (rather than time) basis.

Since variable costs comprise the only changing portion of total cost in the above equation, any change in the aggregate will be equal to the change in variable cost. These changes, due to changes in output, are called *marginal costs.* That is, marginal cost is the change in total cost (equals the change in variable cost) resulting from a unit change in output. Marginal cost is of significance for decisions involving the company's allocation of resources and in product pricing, and is considered further from these practical standpoints later in this and in the following chapters. At the present time, it is sufficient to note that the concept of marginal cost should not be confused with the notion of differential or incremental cost discussed in subsection 5 below.

In economic and accounting theory it is often assumed that variable costs are continuous functions of output when, in reality, some costs that remain fixed over considerable ranges of production increase by jumps, discontinuously, at various levels of output. Costs that exhibit this tendency have been classified as *semivariable* (*semifixed*) *costs.* They consist of a fixed and a variable portion, such as telephone expense, foremen's wages, and certain other expense elements which may remain constant for a wide range of production but then increase by definite jumps as output expands beyond certain levels.

*4. **Short-Run and Long-Run Costs.*** The above distinction between fixed and variable costs bears a close tie-in with another kind of cost dimension used by economists, short-run costs and long-run costs.

Short-run costs are costs that can vary with the degree of utilization of plant and other fixed factors, i.e., vary with output, but not with plant capacity. The shortrun is thus a period in which fixed costs remain unchanged but variable costs can fluctuate with output; in short, it is a period in which a flow of output emerges from a fixed stock of resources.

Long-run costs, in contrast, are costs that can vary with the size of plant and with other facilities normally regarded as fixed in the short run. The latter is an interval of time in which plant, equipment, labor force, etc., can be expanded or contracted to meet demand requirements. It is thus a period in which the firm's output emanates from a variable stock of resources and, therefore, a period in which there are no fixed costs, i.e., all costs are variable. These concepts are discussed further in the following subsection.

The distinctions between fixed and variable costs and between short- and long-run costs are useful for predicting the effect of temporary and permanent output decisions on costs, prices, and profits. Some evidence of this has already been seen earlier in Chapter 4 in the discussion of break-even charts. Further aspects of this will be examined in greater detail later in this chapter where cost functions and their empirical measurement are taken under consideration.

*5. **Differential (Incremental) Costs and Residual Costs.*** When a decision has to be made involving a change in the volume of business, the difference in cost between the two policies may be considered to be the cost really incurred due to the change in business activity. This change in cost is the *differential cost* (also called "incremental cost") of a given amount of business. It represents the change in costs resulting from a change in business activity, where the latter may include any type of change such as the introduction of new machinery, development of a new product, or expansion into different markets. In estimating costs for this purpose, it is necessary to include any interest charge that is actually incurred in the one case and could be avoided in the other case. Thus, cost in this sense depends as much on one of the proposed alternative policies as it does on the other. The cost of remaining in business and producing a product is one thing if the alternative is to go out of business entirely, or it may be another if the alternative is to keep a skeleton force on hand even if the plant is not running at all.[2]

The concept of differential cost forces recognition of the fact that expenses vary according to different dimensions of the business. For example, a trucking company, in considering the taking on of new business, may be confronted with the two alternatives of utilizing (1) more trucks per day, or (2) more payload per truck. The choice of either alternative will result in separate differential costs, and the most economical policy would be to maintain a balance between the two, using each to the point beyond which the other would be cheaper, i.e., where the incremental costs are equal. And in many agricultural processing operations, e.g., meat packing, sugar refining, etc., the differential cost of processing the main product is not the same as the separate differential costs of the various by-products nor of their sum.

---

[2] The concept of differential or incremental cost stems back to Clark, *ibid.*, pp. 49–50. Earlier implications appear in H. J. Davenport, *Economics of Enterprise*, pp. 61–65, 190–91, where it is cast in the same family with opportunity cost.

Evidently, certain costs will not be altered as a result of a decision that will change business activity. Such costs that are not assignable by the differential method are called *residual costs*, and hence are irrelevant as far as the future effects of the decision are concerned. Since differential cost includes interest on investment, i.e., the interest on additional capital that may be required because of the added business, residual cost would include interest on the entire investment to the extent that it is not traceable differentially to some part of the product.[3] It should be noted, however, that differential costs need not be variable with output, nor traceable to a product, nor absolute (cash outlay) costs. They may in many situations be placed in the same family with opportunity or alternative costs (see footnote 2) so that the differential cost becomes the foregone opportunity of using limited resources in their present, as compared to their most profitable alternative, activity.

6. Sunk, Shutdown, and Abandonment Costs. A past cost resulting from a decision that cannot now be revised is called a *sunk cost.* It is usually associated with the commitment of funds to specialized equipment or other facilities not readily adaptable to present or future use, e.g., brewery equipment during prohibition. Sunk costs are thus the costs that result from the permanent and specialized plant. They are important when management is faced with the decision of continuing the existing unit or plant on the one hand or abandoning or replacing it on the other. Beyond this, sunk cost is a minor consideration in any decision affecting the future, and thus contrasts sharply with current-outlay cost which is a present out-of-pocket cost requiring a current cash expenditure.

Shutdown costs may be defined as those costs that would be incurred in the event of a temporary cessation of activities and which could be saved if operations were allowed to continue. The concept is important because of a widely recognized principle that so long as a firm is at least covering its variable costs, it will not cease operations in the short run because any excess will be applied to the recovery of its fixed costs. The principle, though broadly correct, contains certain ramifications. In reality, to suspend operations temporarily involves certain costs that must be considered, such as the compounding and storing of machines, the boarding of windows, and the construction of shelters for exposed property. These are classes of expenses for which interest must also be reckoned as a shutdown cost. Further, additional expenses are incurred when operations are resumed, including the estimated cost of recruiting and training new workers. In essence, the point to be emphasized is that, in the long run, there may be less of a loss if management keeps a few products before consumers, even if revenues fail to cover variable cost, than to close temporarily if shutdown costs (including the costs of re-establishing marketing contacts when business is resumed) turn out to be unduly higher than anticipated.

---

[3] The significance of including interest in calculating differential cost is discussed somewhat briefly in Clark, pp. 50, 65–67; the significance of whether interest as a cost should be incorporated in the firm's books of account or whether it should be treated separately in particular situations when the need arises, is treated extensively in the *Year Book of National Association of Cost Accountants* (1921), pp. 45–96, especially pp. 90–93.

Unlike shutdown costs which are incurred because of a temporary suspension of activities, *abandonment costs* are the costs of retiring a fixed asset from service. The situation may arise, for example, in the abandonment of a war plant or part thereof not useful in peacetime production, of an exhausted mine or oil well, or of trolley facilities upon the institution of bus service. Abandonment thus involves a permanent cessation of activities and creates a problem as to the disposal of assets. Briefly, the correct procedure in such instances is to consider implicit interest on the current market value of the facilities and to depreciate on the basis of sales value. Depreciation based on original cost is manifestly irrelevant for this type of management decision.[4]

*7. **Urgent and Postponable Costs.*** Those costs that must be incurred in order to produce a finished product are classified as *urgent costs.* Examples include the outlays on materials and the expenses of labor in working them up. Costs that may be put off, at least within limits, on the other hand, are called *postponable costs* and include such expenses as maintenance of buildings and machinery. This distinction is significant because, in a certain sense, postponable costs cannot really be postponed. Physical deterioration and obsolescence will reduce the value of a plant whether maintenance is provided for or not. Hence, it is not the cost that is actually postponable, but rather the rate at which it is provided for. Railroads have made fairly common use of the urgent-postponable cost distinction, incurring expenditures for maintenance in periods of low activity in order to repair and maintain equipment that was worn down in periods when activity was high. In the well-managed firm, however, the postponing of such costs is taken out of the discretion of management by recording a regular accrual of depreciation irrespective of whether expenses in a particular period have been large or small. This serves to minimize, though not necessarily eliminate, the postponable outlays, thereby stabilizing costs and, from a social standpoint, reducing cyclical unemployment and fluctuations in purchasing power.[5]

---

[4] These notions of sunk, shutdown, and abandonment costs, and particularly their accounting treatment, are discussed more fully in C. T. Devine, *Cost Accounting and Analysis,* p. 566 and chap. 36, as well as in most other works on cost accounting.

8. Escapable and Inescapable Costs. A cost that may not only be postponed but may be avoided entirely as a result of a contraction of business activity is called an *escapable cost*. It is important to note that such cost is conceived as a *net* figure: the decrease in cost by curtailing or terminating an activity, less any added cost incurred by other operating units as a result thereof. For example, the escapable costs by eliminating a middleman may turn out to be less than originally anticipated if the same functions must be assumed by the selling firm which is less equipped to handle it. Similarly, railroads, for instance, sometimes find it cheaper to retain or perhaps reduce operations on a seemingly unprofitable line than to incur the costs of eliminating it entirely.

An *inescapable cost* (or "unavoidable" cost) is a cost that must be continued in the face of a business retraction. Airlines, for example, must incur certain periodic maintenance expenses irrespective of the volume of business. Manufacturing plants must incur prescribed minimum power costs regardless of the level of sales. Some costs, however, though unavoidable, are nevertheless postponable, as in the case with many types of fixed assets, and frequently costs that appear to be avoided are really only postponed.

Occasionally the escapable-inescapable grouping is employed in place of the more usual fixed-variable classification by some accountants and businessmen. From an economic standpoint, before the enterprise is started and resources are committed, all costs may be viewed as escapable and all expected costs as variable.

9. Controllable and Uncontrollable Costs. This classification of costs is useful mainly as a means for fixing responsibility and measuring efficiency. The distinction may not be too meaningful over the long run, however, for over the life of the enterprise as a whole, all costs are controllable at least in the sense that they are someone's responsibility. Controllability, however, should not be confused with cuttability. Over the long run, not all costs are either equally controllable (e.g., property taxes as compared to materials prices) nor equally cuttable (e.g., skilled union labor as compared to institutional advertising).

10. Replacement Cost and Original Cost. This cost classification has already been discussed more fully in Chapter 4 and hence may be commented upon briefly here. In establishing costs so as to determine income, an asset is conventionally valued on the books at its original cost rather than at the cost of replacing it in the current market. Many accountants and economists have advocated the latter procedure so that, with regard to inventories, the profit figure in the event of substantial price-level changes will represent in a more realistic manner the current situation, as well as improve the results of cost projections as a basis for management decisions.

---

[5] *Cf.* Clark, pp. 55–56. The problem today, however, is perhaps less serious than at the time of Clark's writing (1923). Accountants have come increasingly to accept the practice of recording an expense before the cost is actually incurred. Management, however, still tends to look upon the postponability of expenses as a lifesaver by allowing for retrenchment when times are hard.

When substantial changes occur in the general price level, the firm still has the alternatives of (1) using its materials to produce and sell a finished product, or (2) disposing of the materials at current prices. The sacrifice is thus measured by the market price of the materials and not by their original cost. The difference between the two, assuming that the price has fallen, is a loss due to holding goods during the period, and this loss should not be charged as a cost of making the materials into finished goods. More than one firm has been known to lose contract bids because it figured price on the basis of original cost of materials after the market price had dropped significantly. Management refused to make bids low enough to secure orders because the lower price would not cover certain costs incurred in the past. Actually, the determining criterion should be based on *what the costs will be at the time the order is filled, compared to what they will be at that time if the order is not taken.* It may be easier to charge materials at their original cost, but it is more accurate to charge them at the market price prevailing at the time they are used, thereby separating gains and losses arising from production from gains and losses due to changes in the value of materials in stock. Original costs are bygones and should be re-examined in the light of present market values if it appears likely that such re-examination would give rise to a different decision with respect to pricing.[6]

Conclusion

The foregoing classification of cost concepts reveals various distinctions from both the economic and accounting standpoints.

> This being the case, the thing to do is to cease trying to make one concept do the work of several. After all, the obligations a corporation must meet before dividends are paid are one thing, and the whole financial outgo or sacrifice attributable to the act of producing certain goods is another thing, and a conservative standard for valuing unsold goods is still a different thing. Undoubtedly the ultimate solution lies in the development of systems of cost analysis which shall be separate from the formal books of account, though based on the same data. This analysis will be free to study differential cost and cost as a normal supply-price, without being tied down by the rules that are legitimate and necessary in financial accounting.

---

[6] See Clark, pp. 197–98. It should be observed that the above discussion is using replacement cost in the last-in, first-out sense.

> The economist also uses both these conceptions because they represent forces governing price. One is a long-run standard, the other a natural minimum limit on short-run fluctuations. Abandonment costs play an important part in determining who the typical marginal producer is, and shutdown costs furnish an incentive to maintain production in off times, and a measure of the waste of unemployed capital, so far as it is borne by the business enterprise.
>
> The statistician, as such, has no characteristic purposes, but he has a technique peculiarly adapted to the study of differential costs. The engineer has to deal with the total cost (in the economist's sense) involved in new enterprises, and with comparisons of total cost for different kinds of plant, or for different policies in a going concern where some change of the plant is involved. He must therefore take account of interest on investment. Thus he deals with total cost and with differential cost, but the canons of general accounting are foreign to his needs.[7]

There thus cannot be found any single meaning for "cost of production" that would be universally applicable in all situations. At best, the analyst can only attempt to translate the many current usages into consistent language in order to be certain that the given concept is being used for its proper purpose.

Cost-Output Functions

Before proceeding to the empirical derivation of cost relationships in the following section, it is well to review the nature of these relationships as they exist in economic theory. The following treatment, though far short of being exhaustive, provides a brief sketch of the essential nature of cost-output functions both under short- and long-run conditions. In both cases the relevant cost concepts involved are the ones commonly employed by economists, namely the fixed-variable classification and the subdivision of marginal costs, all of which were discussed above. These costs apply even in the most simple production processes and are the ones that are derived by econometricians when they construct cost curves from actual company data.

Short-Run Costs. In production economics the *short run* is defined as a period long enough to vary output by altering the combination of variable to fixed factors. The short run thus refers to a cost structure and time period in which some of the productive factors (e.g., plant, basic equipment, management) are fixed in quantity and form. A firm with a given arrangement of productive factors will experience one short-run cost situation, while another firm twice as large with twice as many resources available will have a different short-run cost curve. The *long run* refers to the cost structure of a firm over a period of time long enough so that no factors need be considered as fixed, or in other words a period of time long enough so that all of the firm's costs are variable. In the following paragraphs *short-run* costs are discussed first, including their more important ramifications; the nature of *long-run* costs is then outlined so as to unify the logic into a consistent body of doctrine.

---

[7] Clark, pp. 68–69.

The level of the various cost curves—fixed, variable, total, and marginal—will be affected by factor prices, but the exact nature (curvature) of the curves will depend on the nature of the underlying production function as discussed in the previous chapter. The fundamental starting point in the development of cost theory is that a unique functional relationship exists between cost and the rate of output for a firm. Admittedly, there may be independent variables other than output that will affect cost (e.g., lot size, factor prices, etc.), but these are assumed to remain constant in constructing the cost curves. The curves thus derived are static or timeless in nature, meaning that they show only the various costs that will prevail under alternative output levels. This static characteristic, it will be recalled from basic economics, is also applicable to the theoretical revenue (including demand) curves, in that they reveal the various receipts of the firm for alternative purchase quantities on the part of buyers. When cost or revenue functions are derived from observations over time, on the other hand, they are not the same static functions of economic theory but rather an expression of the average relationship over time. But more of this in the following section. At the present our interest turns to the nature of the theoretical cost curves as determined by their underlying production functions.

FIGURE 7-1

THREE KINDS OF SHORT-RUN TOTAL COST FUNCTIONS

A, Constant Productivity; *B*, Decreasing Productivity; *C*, Increasing Productivity

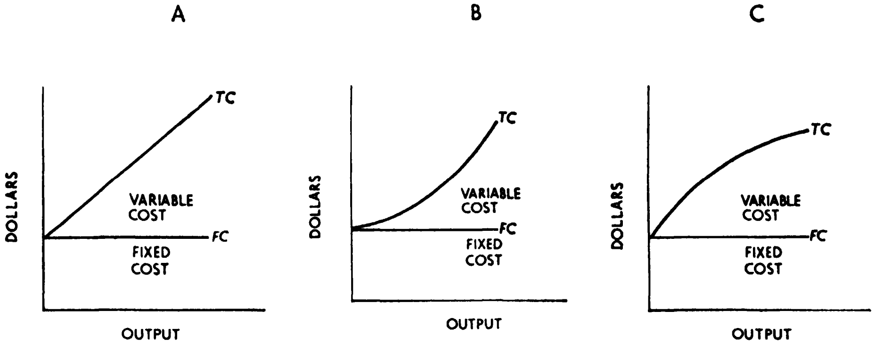


Figure 7–1 illustrates three kinds of short-run total cost curves under conditions of constant, decreasing, and increasing productivity, with output scaled horizontally and dollar costs vertically. In Figure 7–1A the *TC* curve is linear, thus indicating a constant price per unit of variable input purchased, and hence the fact that each unit of input as well as each unit of output adds the same amount to total cost. This type of linear cost function exists over a range of output when, assuming constant price levels

and technology, the fixed factors are readily divisible so that the fixed and variable resources can be mixed at minimum-cost proportions for each output level (as discussed earlier in Chapter 6). At zero output, fixed cost (cash rent, taxes, insurance, and depreciation as a function of time and obsolescence) equals total cost, while at higher output levels the difference is represented by variable costs. The total cost curve, it may be noted, would turn sharply upward for output levels beyond the physical capacity of the equipment.

Figure 7–1B shows a total cost function where the factor-product relationship is one of diminishing marginal productivity throughout the entire range of output. The reason for this is that even if each unit of the variable factor costs as much as any previous unit, each additional unit of the input adds less to total output than the previous unit. (The elasticity of the production function is less than 1 throughout.) This illustration of diminishing returns throughout the entire output range occurs when the fixed factor is limited and not divisible. The shape or curvature of the *TC* curve is due solely to the technical nature of the input-output relationship and not to market conditions or factor prices.

FIGURE 7–2

GENERALIZED SHORT-RUN TOTAL COST FUNCTION

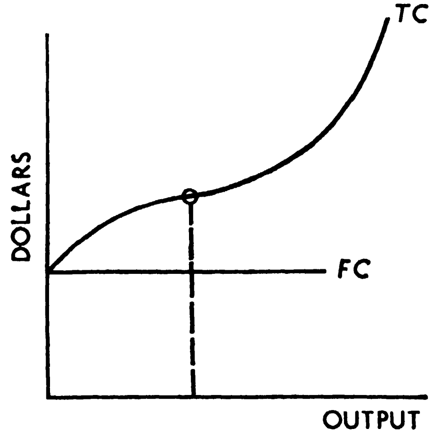


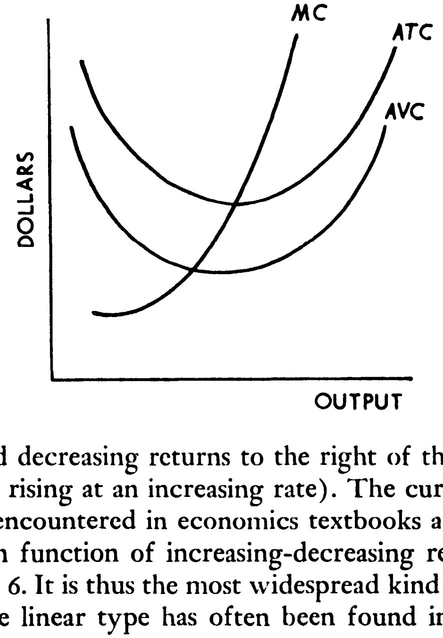
Figure 7–1C shows the total cost curve for a production process under conditions of increasing returns throughout the entire output range. This means that each unit of output adds less to total cost than the previous unit, and this in turn is due to the fact that each unit of input in the underlying production function (whose elasticity is greater than 1 throughout) adds more output than does the previous unit of input. Actually, the possibility of an enterprise experiencing increasing returns for all output levels is unlikely; at best, it may perhaps be found at the lower levels of production where the fixed factors are excessive relative to the variable, and before the stage of diminishing productivity sets in. Incidentally, it

should be noted, as must be evident by now, that the curvature of the production function and the total cost curve is always reversed in that when one is concave, the other is convex, and vice versa, except with a linear relationship.

The most common cost functions are those that combine the phases of both increasing and decreasing returns. Most cost functions that appear to be of a constant or increasing returns nature are more likely to be only segments of curves and, if they could be extrapolated, would eventually exhibit a phase of decreasing returns. Figure 7–2 illustrates this generalized type of cost function with increasing returns resulting at all levels of output to the left of the vertical dashed line (because total cost rises at a decreasing rate) and decreasing returns to the right of the dashed line (because total cost is rising at an increasing rate). The curve as shown is the type commonly encountered in economics textbooks and is based on the classic production function of increasing-decreasing returns such as discussed in Chapter 6. It is thus the most widespread kind of total cost function, although the linear type has often been found in empirical studies for reasons to be discussed later in this chapter.

FIGURE 7–3

GENERALIZED SHORT-RUN AVERAGE AND MARGINAL COST FUNCTIONS



To improve comprehension of a firm's cost structure as well as serve as a better basis for various kinds of decision problems confronting management, the average and marginal cost curves are necessary. For most purposes these include average total cost (*ATC*), average variable cost (*AVC*) and marginal cost (*MC*). All of these can be derived from the total cost data. Thus, $ATC = TC \div \text{output}$; $AVC = TVC \div \text{output}$; $MC = \Delta TC \div \Delta\text{output}$. Numerous other methods can be employed in deriving these curves from given output and total cost data, since $TC = FC + VC$ is the basic relationship, and the quantities can be algebraically transposed as desired. In Figure 7–3 the *ATC*, *AVC*, and *MC* curves corresponding to the generalized total cost curve of Figure 7–2 are presented. Note that the *MC* curve passes through the minimum *ATC* and *AVC* points in accordance with the rule of the "average-marginal relationship" explained earlier in Chapter 6. Other than this, the various curve relationships should already be familiar to the reader from his elementary education in economics and hence require no further discussion at this time. The only point that need be mentioned is that, as with the total cost curve, the shape (curvature) of the average and marginal curves is conditioned by the technical nature of the underlying production function and not by factor prices. A change in the latter will shift the curves up or down but will not affect the slopes as such.

Long-Run Costs. The analysis of short-run costs reveals how a firm's costs will vary in response to output changes within the limits of a time period short enough so that the size of the plant may be regarded as fixed. By extending the logic one step further, it is possible to develop a *long-run cost curve* or function which, correspondingly, is one that shows the variation of cost with output for a period long enough so that all productive factors, including plant and equipment, are freely variable in amount. The knowledge of such a long-run cost curve, or "planning curve" as it is also called, can be of use to management in determining output rates over periods long enough so that assets acquired for use during the period can be fully amortized, and in establishing rational policies as to optimum plant size, location, and operational standards in general.

The formal nature of the relationship between long-run costs, short-run costs, and size of plant may be established conceptually in this way. At the planning stage when management is considering the erection of a plant, it is faced with the problem of selecting one of many alternative combinations of fixed and variable factors. But at the planning stage all factors are variable, and only for each "output level" will there be a given production function and cost structure. Assuming that production costs and the nature of market demand were known for each particular layout, the appropriate layout could be determined. The long-run average cost curve, *LAC* in Figure 7–4, would then show for each possible output level, the lowest cost of producing that output, assuming plant size and intensity of utilization are covaried to obtain the best results. The following principles are thus evident or can be deduced.

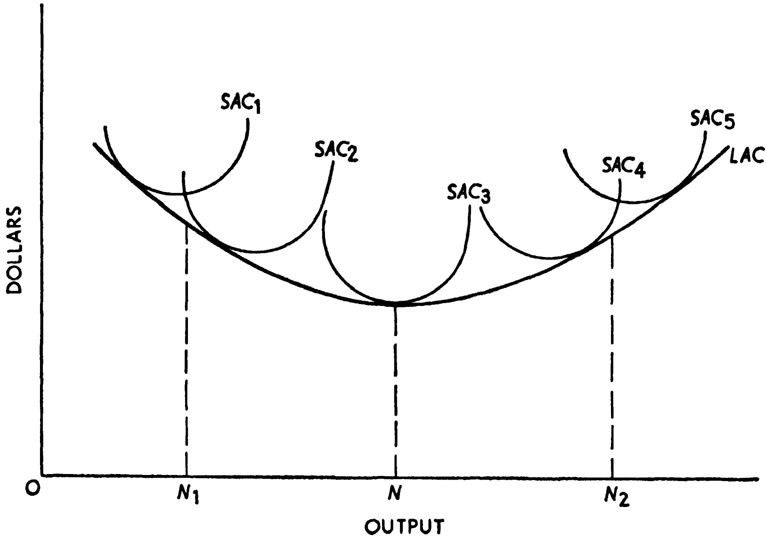
1. There will exist a different short-run average cost curve, *SAC*, for each possible plant size or for each technique of production (production function). There is thus an entire family of short-run cost curves, each corresponding to a particular point on the long-run average cost curve. That is, although only five *SAC* curves are shown, infinitely more could be drawn, depending on the divisibility of productive units and their technical nature.

2. The *LAC* curve generalizes the entire family of *SAC* curves by enveloping them together. The U shape of the long-run curve implies at first lower and lower average costs until the "optimum" scale of plant shown by SAC_3 is reached, and thereafter successively higher average costs with larger plants.

3. The *LAC* curve is tangent to only one point on each *SAC* curve as in Figure 7–4. The tangency point occurs (*a*) to the left of the minimum-cost point on all short-run curves, which in turn are to the left of the optimum curve SAC_3, and (*b*) to the right of the lowest cost point on all short-run curves that are to the right of the optimum curve.

FIGURE 7–4

Short- and Long-Run Average Cost Functions



For the optimum curve, however, the tangency occurs at the minimum point on that curve, i.e., at the lowest point on SAC_3. Therefore, for outputs less than *ON* for which the optimum scale is SAC_3, it is more economical to "underuse" a slightly larger plant operating at less than its minimum-cost output than to "overuse" a smaller plant. For example, it would be cheaper to produce output ON_1 with a plant designated by SAC_2 than with one represented by SAC_1. Conversely, at outputs beyond the optimum level *ON*, it is more economical to "overuse" a slightly smaller plant than to "underuse" a slightly larger one. Thus, it is cheaper to produce ON_2 units with plant SAC_4 than with SAC_5.

5. Finally, the tendency for long-run average costs to fall as the firm expands its scale of operations is a reflection of cost economies that are frequently encountered with increasing size, while the ultimate rise in the long-run curve is due largely to the eventual setting in of disecon-

omies of large-scale management. The latter was discussed earlier in Chapter 1 in the section on "Sequential Decisions" where it was pointed out that as the firm becomes larger and decision making more complex, the burden of administration becomes disproportionately great and "diminishing returns" to management set in.

Conclusion

The above discussion of cost theory could only touch upon some of the barest essentials. Nevertheless, enough has been said to provide the necessary background for comprehending the measurement of cost relationships as a basis for improved decision making by management. It should be borne in mind that cost-output relations are *anticipated* relations drawn, frequently, from past experience. As such, they are subject to uncertainty and it can only be hoped that their measurement will be of some aid in reducing the degree of uncertainty inherent in a decision problem. Where the uncertainty is great, management's knowledge of strategic costs may at best be sketchy, and some measure of important relationships is better than none at all. On this basis, we turn our attention now to the area of cost measurement with particular emphasis on the findings of some important studies in that field.

COST MEASUREMENT

In this section we are concerned with the construction of empirical cost functions or, in other words, the measurement of the actual cost-output relation for a particular firm or group of firms. Several methods exist by which an analysis of costs can be undertaken, but three of these, the accounting, engineering, and econometric approaches, are the most common.

Accounting Method. Essentially, the method used by cost accountants is to classify the data into various cost categories (e.g., fixed, variable, semivariable) as described earlier, and then to take observations at the extreme and various intermediate output levels. In this manner linear or curvilinear cost functions are estimated and built up from the basic data, with little or no attention normally paid to changes in factor prices or other conditions that may have affected costs.

Engineering Method. In the engineering method, emphasis is placed primarily on the nature of physical relationships such as pounds of supplies and materials used, rated capacity, etc., and these relationships are then converted into dollars to arrive at an estimate of costs.[8] The method may be particularly useful when good historical data are difficult to obtain and the analysis may therefore require a relatively greater utilization of engineering rather than economic theory.

---

[8] The basic procedure is simply illustrated in W. Rautenstrauch and R. Villers, *The Economics of Industrial Management*, chaps. 12 and 13, although in recent years more refined techniques have been developed.

Econometric Method. The econometric or statistical approach uses statistical tools, such as correlation analysis, combined with economic theory, to measure the net effect of output variations on cost. The goal is to construct a cost curve or function from historical data that will reflect as closely as possible the static cost curve of economic theory. As stated earlier, however, since the empirical curve is at best only an average of past relationships, it is not an exact replica of the theoretical cost curves discussed in economics textbooks.

These three methods of cost measurement should not be regarded as competitive, but rather as supplementary and complementary to one another. As always, the choice of any method depends on the purpose of the investigation, i.e., what it is that management really needs and wants, and the time and expense considerations in selecting one method over another. From the standpoint of this book, however, it is the econometric method that is of primary interest, and hence the one that will occupy our attention throughout most of this and later sections.

Analytical Framework for Cost Measurement

In the first section of Chapter 5, dealing with the measurement of demand, certain measurement problems were discussed concerning various adjustments in the data and other considerations that must be taken into account when attempting to derive empirical functions from economic data. As in the measurement of demand, so too in the measurement of cost, there are problem areas of a methodological nature to be considered. Since several of these difficulties have already been discussed or implied elsewhere, some of the more essential ones may be sketched briefly at this time.

Basically, the problem is to derive a cost function, expressed mathematically as an equation or geometrically as a curve, that will show the net relationship between the firm's costs and its rate of output. If the shape of the cost curve depended solely on the rate of output, solving the problem would be fairly simple. Unfortunately, costs depend on a number of factors in addition to output, so the problem resolves itself into eliminating these other cost determinants in order to arrive at a cost function that reasonably expresses the cost-output relation. Generally speaking, the following problem areas must be handled in preparing the empirical analysis.

Time Period. The choice of an appropriate time period on which to base the analysis involves three important considerations: normality, variety, and length of observation.

Normality. The time period should be a "normal" or typical one for the firm, so far as this is possible. This means that the period covered should be one which was reasonably static in that changes in technology, plant size, efficiency, and other dynamic occurrences that may have a significant bearing on costs were either nonexistent or at least at a minimum. Admittedly, a completely static period would probably be impossible to find. However, a period in which changes were relatively minor would be acceptable if the data could be adjusted to compensate for the differences; if not, the cost function will not reflect the typical type of cost behavior desired.

Variety. The period should be one in which there were sufficiently wide variations in output so that enough observations for a correlation analysis can be obtained. Further, since the results are to be used as a guide for future planning, the period should be recent enough to include data that will be basically relevant for the future. In many instances a minimum of three to five years is used as a source for the data within a period of a business cycle, say seven to ten years or thereabouts. On the other hand, if the normality conditions stated above are satisfied, a full business cycle may be preferable as an analysis period.

Length of Observation. The period chosen should be one in which the observational unit (week, month, quarter, or year) will be a minimum to the extent that completeness of the data will permit. A small observation unit such as a week or perhaps a month will allow measurement of slight output variations more readily, say, than will a year. Further, the cause-effect relationship between output and cost is more readily discernible with small rather than large observational units. Frequently, if not usually, the month is the most typical unit chosen in cost studies, although in the steel studies discussed below the analyses were conducted on both a quarterly and annual basis because of technicalities involving inventory changes and cost reporting dates.

Technical Homogeneity. In order to minimize the effect on costs of differences in product, equipment, frequency of production lags, etc., the plant chosen for a statistical cost study should be characterized by an input and output structure that is as technically homogeneous as possible. This means that on the input side, the use of identical or very similar units within factor classes is necessary, e.g., equipment, so as to prevent variations in cost due to different machines being brought into production at different output levels. On the output side it means that the different products produced should, ideally, be small in number so as to facilitate measurement, and that they not undergo significant cost changes (due, for example, to changes in composition or style) during the analysis period. If these conditions of output homogeneity are not met, the analysis may require that a weighted index of output be constructed for products or classes of products according to some logical criterion. Various approaches are possible just as in the construction of demand curves discussed in earlier chapters. Tons shipped rather than tons produced may be more useful, for example, if inventory fluctuations (the difference between production and shipments) are relatively small, as in one of the steel studies discussed later. In a cost study made for a clock manufacturer, the weights used for the output index were based on direct labor costs; in a cost analysis of a men's clothing factory, on the other hand, square feet of wool of a specific grade was chosen as the measure of output, from which conversion coefficients were derived so as to apply to other types of materials used by the company. In short, a number of preliminary measures must frequently be developed based on theoretical considerations, and then the particular one chosen is the one that seems to fit the data best.

Cost Adjustments. The third problem area in cost analysis involves decisions as to the proper choice of data and the types of adjustments needed to correct the figures if they are to be recast into a meaningful cost function. The problem as a whole breaks down into three subclassifications: cost inclusion, deflation, and cost-output timing.

Cost Inclusion. Since the object is to arrive at a cost-output relation, the problem is to select only those elements of cost that vary with (are functionally related to) output. Thus, various kinds of overhead and allocated expenses that do not bear any relation to production rates should be excluded. Sometimes a series of preliminary correlations must be made to determine which costs should and should not be included in the final analysis. Also, it should be mentioned that total rather than unit (average) costs should be used in conducting most econometric cost analyses, for two main reasons: (1) the results are likely to be more reliable statistically because average cost is a ratio of two variables and therefore more susceptible to error, which in turn may cause magnified errors in the derived marginal cost function; (2) the marginal and average cost functions can be readily derived mathematically from the total cost function if desired, or even by simple arithmetic if a cost table or schedule is constructed from the total cost equation (analogous to the production function schedule constructed in Chapter 6, Table 6–1 for the auto laundry study).

Deflation. As in the construction of empirical demand functions, so too with cost functions, the data must usually be reduced or deflated to a particular base period if the results are to be meaningful. Wages and equipment price indexes are readily available and are frequently used for such purposes, or the analyst may construct his own indexes if it seems desirable. In any event, the purpose of deflating the data is to adjust for significant changes in input prices during the analysis period, as has occurred for example with virtually all types of productive factors since World War II.

Cost-Output Timing. The third area of an adjustment nature involves the problem of obtaining the correct correspondence of cost and output. Costs are not normally recorded in the books of account in such a manner that they are readily traceable to the output variations that created them. Usually, technical engineering estimates will be necessary if the correct timing associations are to be established between the two variables, cost and output. Of particular importance in this respect are certain costs that are usually charged as a function of time, such as depreciation (normally on a "straight-line" basis). These costs, or portions thereof, must first be adjusted or recalculated as a function of output rate before they can be incorporated in the over-all cost function.

Choice of Equation or Curve. Finally, as in the derivation of demand functions, there is the problem of choosing the type equation or curve which seems to fit the data best, justified as far as possible by economic theory, before the correlation analysis can be made. Referring back to the diagrams, the total cost curve may be either linear as in Figure 7–1A, in which case the form of the function would be

$$Y = a + bX\,;$$

it may have a bend as in Figures 7–1B or 7–1C for which the equation form is

$$Y = a + bX + cX^2\,;$$

or it may have two bends as in Figure 7–2 and thus be a cubic function of the form

$$Y = a + bX + cX^2 + dX\ .$$

In all three instances Y represents total cost, X is output, and the letters a, b, c are constants whose probable values are to be determined by correlation analysis. In a great many if not most of the empirical cost functions that have been derived by econometricians, the results were linear, which means that the total cost curve was a straight line and therefore marginal costs were constant over the range of output considered. It is likely, however, that at some production level beyond the range of the data, the total cost curve would definitely bend upward and the marginal cost curve would also eventually rise. In the discussion that follows, cost functions representing both linear and curvilinear relationships are illustrated, the latter of the second-degree parabola form, as in the second equation above. Cubic or third-degree parabolas, as in the third equation, however, are not illustrated empirically because they seem to represent only a theoretical generalization. In practice, attempts to fit cubic functions have not ordinarily yielded statistically significant results because of the difficulty of distinguishing actual discontinuities in the total cost curve from the random scatter of the observations.

Long-Run Costs

The measurement framework discussed above is applicable primarily to the derivation of short-run cost functions, but some of the same considerations and others too apply to the measurement of long-run costs as well. Basically, these other difficulties break down into three classes of problems: choice of method, measurement of size, and measurement of cost.

Choice of Method. There are two choices open to the analyst who wants to derive a long-run economic cost function: (1) he can analyze changes in the same plant's costs at different points in time, or (2) he can analyze changes in costs of different size plants at the same point in time. The first approach requires, among other things, a virtual constancy of such dynamic factors as technology and product line to say the least, or else the data will probably be impossible to adjust. From a practical standpoint, therefore, the method could be considered as a possible approach only for a plant that has remained relatively static for long periods of time. The second approach is usually the preferred one for these reasons, but it also raises several difficulties, particularly where the industry is not sufficiently homogeneous so that differences in accounting methods, management, technology, etc., tend to cloud the relationships of cost to plant size. If the various plants are owned by a single firm, some of these obstacles are avoided, but this becomes a special rather than a general case. In the following section the findings of an empirical study involving a number of plants under separate ownerships is considered for this and other reasons to be explained later.

Measurement of Size. Some measure of size of plant that accords with theoretical considerations is necessary if an empirical long-run cost curve is to be meaningful. Some typical physical measures of size include rated capacity, number of workers, man-hours, and machine-hours; economic measures include various balance sheet items such as total assets or net worth, and "normal" or "average" output expressed, perhaps, as a per cent of capacity. Unfortunately, there is no simple solution to the problem of choosing an appropriate measure of size. This is because most of the various physical measures require certain conditions of homogeneity within and between plants if the firms are to be ranked in size, and because the economic measures do not meet the theoretical requirement that the long-run curve be an envelope of the various short-run curves, i.e., the latter will not produce the correct long-run curve unless a correct measure of economic capacity is established. However, to the extent that an engineering measure of capacity is roughly equal to minimum average cost, which is an approximate concept of economic capacity, it may be possible to average statistically the various cost-output observations in industries with a wide array of plant sizes in order to arrive at a meaningful function whose curvature approximates theoretical requirements.

Measurement of Cost. Finally, as with short-run cost functions, there is the problem of removing those factors that are irrelevant to the analysis so that a cost–plant size function can be obtained. Some of these "other" variables to be removed or accounted for are differences in accounting procedures, changes in factor prices, locational differences, product differences, and differences in output rates. If the effects of these variables are not removed, they will affect average costs and thus conceal the true relationship between cost and size of plant. To some extent the various adjustments discussed thus far will serve to eliminate some if not most of these differences in many instances. Other than this, no general set of rules or procedures can be laid down that will be universally applicable in all cases. Each cost–plant size study involves a different set of circumstances to be handled separately, and at best only illustrative models can be presented (as done below) from which the analyst must formulate his own critical evaluations.

Short-Run Costs: U.S. Steel

We summarize below the results of two outstanding econometric cost studies of the United States Steel Corporation. The first was done under the supervision of T. O. Yntema in late 1939 in connection with the Congressional investigation at that time, and is referred to hereafter as the "Yntema study."[9] The second, done independently of the first, was prepared at approximately the same time by Kathryn Wylie and Mordecai Ezekiel, and will be referred to here as the "Wylie-Ezekiel study."[10] The procedure followed will be to discuss first the separate findings of these studies and then to present a comparative analysis of the two. The results are particularly interesting not only from a methodological standpoint, but for their practical value to management, as will be seen shortly.

Yntema Study. The short-run total cost curve derived in this steel study was a composite curve for the Corporation and its subsidiaries. Since different adjustments had to be made for different types of costs, the total costs of the Corporation and its subsidiaries (exclusive of intercompany items and costs of extraneous nonoperating transactions) were broken down for each of the years 1927 through 1938 into seven categories: (1) interest, (2) pensions, (3) depreciation and depletion, (4) taxes other than social security and federal income and profits taxes, (5) payroll, (6) social security taxes, and (7) other expenses. Each of these costs as they existed in past years was separately adjusted to 1938 levels or treated in the following manner:

1. *Interest* cost, since it is not related to volume, was converted to 1938 conditions by substituting the 1938 interest charge in the figures for each year.
2. *Pensions* cost, like interest, was converted to 1938 conditions by substituting the 1938 figure.
3. *Depreciation and depletion* costs were not adjusted because there had been no significant change in the Corporation's accounting procedures during the period.
4. *Taxes other than social security and federal income and profits taxes* were adjusted for the changed tax laws (which followed substan-

---

[9] *Hearings* before the T.N.E.C., Part 26, *Iron and Steel Industry*, Exhibit No. 1417, 1940.

[10] "The Cost Curve for Steel Production," *Journal of Political Economy* (December, 1940).

tially the same pattern after 1932 but allowed for considerably lower taxes prior to that year) by substituting in prior years the taxes for the volume involved that were indicated by the 1932–38 regression line between taxes and volume.

5. *Payroll* costs for each year of the analysis period were adjusted to 1938 rates according to the proportionate change in average hourly earnings between the year in which the payroll was incurred and 1938.

6. *Social security taxes* at 1938 rates for the various amounts of payroll were estimated by applying the 1938 ratio of these taxes to payroll.

7. *Other expenses,* consisting largely of goods and services purchased by the corporation from others, were adjusted by the BLS wholesale price index for commodities other than food and farm products. Material costs, which usually fluctuate in accordance with operating rates, were adjusted to 1938 price levels in order to ascertain the changes in unit costs due to changes in volume alone.

Federal income taxes were omitted from the analysis because they are a function of profit rather than volume, and it was the cost-volume relationship with which the study was concerned. In general, the total costs obtained by adding together the adjusted items for each year represented what the costs would have been if 1938 prices, wages, interest and tax rates had prevailed. A further adjustment was then made, based on the downward cost-volume trend during the period, to account for the extent to which the same tonnages could have been produced in 1938, due to increased efficiency, at lower cost than they could have been produced in prior years at 1938 prices, wages, pensions, interest, and tax rates.

The final adjusted costs and the related weighted tonnages of shipments resulted in the linear cost curve shown in Figure 7–5. Letting Y represent the adjusted cost data on the vertical axis and X the amount of steel, in millions of weighted tons shipped, on the horizontal axis, the equation of relationship is

$$Y = 182.1 + 55.73X.$$

The equation was derived by simple least-squares regression methods.[11] It reveals that $182.1 million represents fixed costs which remained constant irrespective of output volume during the analysis period, while

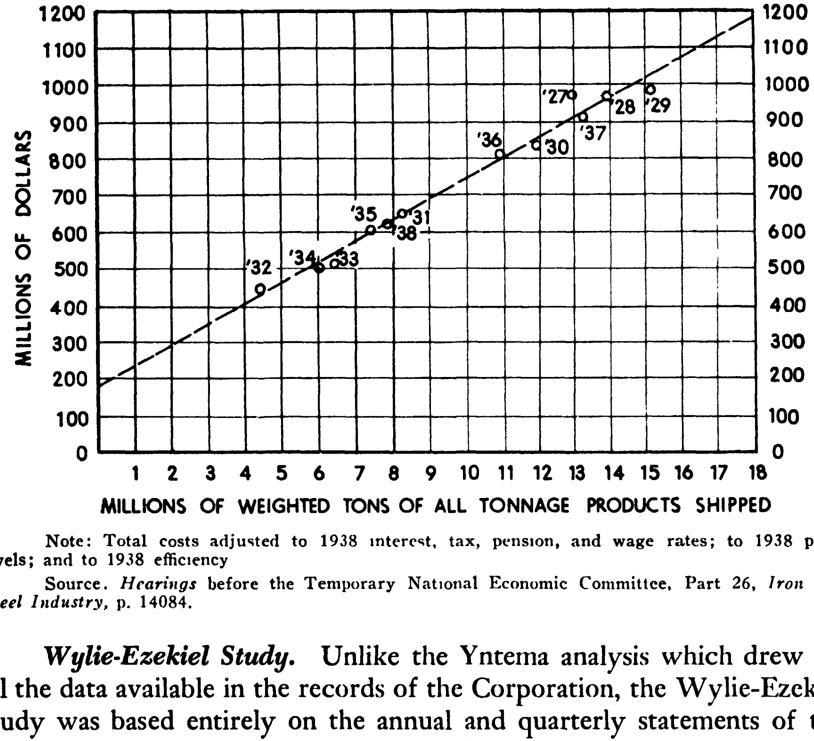
---

[11] The classical method of least squares raises some questions of doubt as to the validity of the derived estimates, not only in this study but in similar ones not treated here, e.g., Joel Dean's well-known studies of a leather belt shop, a hosiery mill, and a department store. In brief, these analysts frequently have neglected (1) the possibility that the data may not be randomly distributed, i.e., that the consecutive observations may not be independent, (2) the consideration that there are errors in the variables, and (3) the fact that the cost function is only one of a system of simultaneous relationships. All of these factors, however, have been brought out in earlier chapters and need not be discussed again. They are stated here only to point out that many of the problems relating to demand measurement apply to production and costs as well (see Chapters 3 and 5).

$55.73 represents the additional or marginal cost of each weighted ton of product shipped. Since the total cost curve is linear, the marginal cost curve, of course, is flat, but the average total cost curve will decline with increased volume. While additional costs might possibly vary if output were pushed beyond either of the extreme output limits shown in the diagram, it should be noted nevertheless that the study, based on the 1927–38 experience, includes annual rates of operation ranging from approximately 18 to 90 per cent of ingot capacity.

FIGURE 7-5

RELATIONSHIP BETWEEN TOTAL COSTS OF OPERATION AND VOLUME OF BUSINESS
1938 Conditions, U.S. Steel Corporation and Subsidiaries



Note: Total costs adjusted to 1938 interest, tax, pension, and wage rates; to 1938 price levels; and to 1938 efficiency

Source. *Hearings* before the Temporary National Economic Committee, Part 26, *Iron and Steel Industry*, p. 14084.

Wylie-Ezekiel Study. Unlike the Yntema analysis which drew on all the data available in the records of the Corporation, the Wylie-Ezekiel study was based entirely on the annual and quarterly statements of the Corporation and the general statistics of the steel industry. Yet, even with the more accurate data, the results secured by the Corporation were similar to those secured by Wylie and Ezekiel. The results of the latter are interesting, therefore, as a technique which may be useful for other firms and industries where detailed internal data are not available.[12]

---

[12] In contrast with the previous footnote, it should be stated that in the present study the statistical significance of the available series was weakened by two factors: (1) the high degree of serial correlation between successive observations, and (2) the high intercorrelation among many of the independent variables. To reduce the effects of serial correlation on the true correlation, annual as well as quarterly series were

The published results of the study consisted of two quarterly analyses—one covering the period 1929 through 1937, and one 1932 through 1937—and two annual analyses covering 1920 through 1934. In general, the procedure was to analyze separately the relation of depreciation and depletion to sales for which linear relationships were found, and then the relation of direct costs to such variables as percentage of capacity operated, average hourly earnings, the price of steel scrap, and labor efficiency. It was found that approximately 85 per cent of the variation in direct costs per ton of finished steel produced was accounted for by the independent factors used. Adding the cost estimates obtained from the analysis of depreciation and depletion and those from the analyses of direct costs, total annual costs were obtained. In order to arrive at a composite cost curve from the analysis showing the relation between total cost and percentage of capacity operated, with other factors held constant, the estimates of depreciation, wage costs, and other costs were added together for whatever constant values of the other variables were assumed. The results obtained from the two annual analyses are shown by the two broken-line cost curves in Figure 7–6.

Comparative Analysis.[13] Both the Yntema study and the Wylie-Ezekiel study were made independently of one another and both were concerned with the empirical derivation of cost functions for the United States Steel Corporation. Both studies used the same commercial concept of cost—that used by the Corporation in its accounting records. Attempts to separate the prime costs and indirect costs of economic theory were not made, probably because it would have been impossible to do so from the published records available. The Yntema analysis was based on actual charges made by the Corporation in its accounts, while the Wylie-Ezekiel study attempted to approximate those charges as closely as possible to arrive at meaningful estimates. In the Yntema study, the cost items for different years were adjusted to 1938 conditions and a separate analysis was made of each item of cost as a function of the weighted average production each year from 1927 through 1938. These estimated costs were then added together to obtain a composite cost function. In the Wylie-Ezekiel study, depreciation and depletion were analyzed separately, but remaining costs were converted to a cost-per-unit basis which then became the dependent variable in a multiple correlation analysis. Cost per ton was related to per cent of capacity operated (as reported by the Corporation), to wage rates, to prices of steel scrap, to labor efficiency, and so on. The two annual studies (quarterly studies were also made) were

---

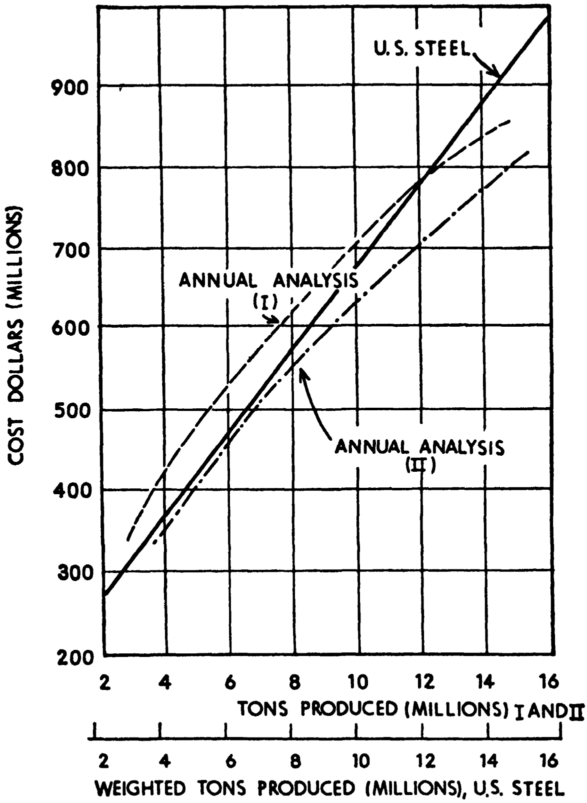
used. The high intercorrelation among the independent variables as well as other difficulties cut down the reliability of the results secured. These limitations should be kept in mind in the present summary of the findings.

[13] See also Wylie and Ezekiel, *op. cit.*, pp. 808–13, and Ezekiel and Wylie, "Cost Functions for the Steel Industry," *Journal of the American Statistical Association*, Vol. 36 (1941), pp. 91–99.

based on the years 1920 through 1934. The data for subsequent years were then used to test the reliability of the results and it was found that in most cases the fit on this extrapolation was reasonably satisfactory. This in brief was the methodology employed in both studies. Now for the results.

FIGURE 7–6

TOTAL COSTS OF STEEL PRODUCTION UNDER 1938 CONDITIONS, EXCLUDING TAXES AND INTEREST



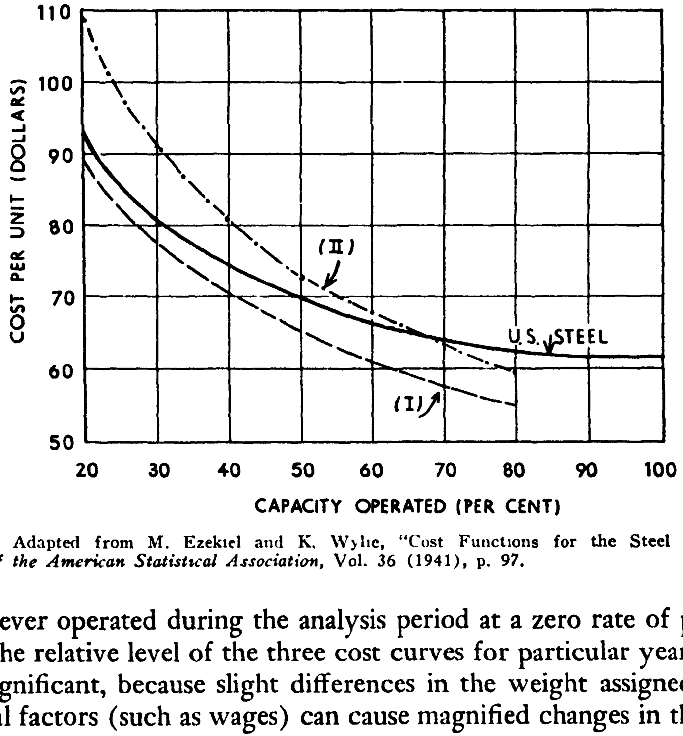
Source: Adapted from M. Ezekiel and K. Wylie, "Cost Functions for the Steel Industry," *Journal of the American Statistical Association*, Vol. 36 (1941), p. 96.

Both studies intended to reveal the general nature of the cost curve for the Corporation, and as it happened they both indicated a similar cost curve. In both studies the relation of depreciation and depletion charges was found to be about the same, but because the two studies were substantially different in method, the results for costs other than depreciation are not directly comparable. In order to compare the two studies, it was necessary to extrapolate the Wylie-Ezekiel analysis to 1938, the base year for the Yntema study, to yield the total cost curves under 1938 conditions. This has been done in Figure 7–6 showing the two curves derived by Wylie and Ezekiel called Annual Analysis I and II, respectively, and the U.S. Steel curve derived by Yntema's staff.

The curve derived by the Yntema study showed a straight-line relationship down to zero output. Since the slope of a straight line is constant, this means that the marginal cost curve would be flat or horizontal over the same output range. The Wylie-Ezekiel study, on the other hand, produced curvilinear total costs that were less steep than the Yntema curve, and were concave to the base (i.e., increasing at a decreasing rate) which means that marginal costs would decline as output increases. Also, they were not extended to estimate cost at zero output because the company never operated during the analysis period at a zero rate of production. The relative level of the three cost curves for particular years is not very significant, because slight differences in the weight assigned to individual factors (such as wages) can cause magnified changes in the level. Thus, the estimate from the Wylie-Ezekiel study was extrapolated four years beyond the analysis period, and 1938 was a year of high wage rates. On the other hand, had the comparison been based on some other year, say 1925 or 1928, the relative position of the curves might have been materially different.

FIGURE 7–7

RELATION OF TOTAL COST PER UNIT OF STEEL TO PERCENTAGE OF CAPACITY OPERATED, 1938 CONDITIONS



Source: Adapted from M. Ezekiel and K. Wylie, "Cost Functions for the Steel Industry," *Journal of the American Statistical Association*, Vol. 36 (1941), p. 97.

The three total cost curves in Figure 7–6 can be converted into average cost per ton as shown in Figure 7–7. All three curves show that costs decline sharply to 50 or 60 per cent of capacity, and above 60 per cent the rate of decline decreases considerably. Both of the Wylie-Ezekiel curves are steeper at higher operating rates than is the Yntema curve, but in the 20 to 50 per cent range of capacity the curve labeled *I* is very similar to the U.S. Steel curve. Probably, a further analysis would be necessary to determine whether the differences in the results of the two studies for outputs above 50 per cent of capacity are significant, and, if so, which study is more nearly correct. Information of this type could be of use to management by helping to indicate the costs that may be expected at various operating levels as well as the profits that may be expected if the required demand and price information are known.

Long-Run Costs: Milk-Processing Plants

To illustrate next the nature of a long-run cost-volume analysis, we summarize here the methodology and results of an interesting study of milk-processing plants.[14]

The purposes of the study were (1) to determine the nature of existing relationships in specialized butter-powder (nonfat dry milk solids) plants between scale of operations and efficiency of use of labor, equipment, and other resources, and (2) to develop cost standards that would be of use to management in making decisions for improving efficiencies of plant operations. Accordingly, twelve butter-powder plants located in Washington, Oregon, and Idaho were studied intensively by gathering detailed information from records, by observation, by physical measurement, and by interview. Both physical and monetary data were obtained for the years 1948 and 1949, and the study was limited to an analysis of processing costs from the moment milk is received at the plant to the loading of the finished butter and powder.[15]

The plants as complete units were heterogeneous due to such characteristics as the diversion of whole milk, cream, and nonfat milk to other plants, seasonality of milk receipts, volume of output in relation to capacity, and various institutional factors. To provide a basis for interplant comparisons, therefore, a functional analysis was made by performing a minute breakdown of the basic production processes necessary to the production of butter and nonfat dry milk. This breakdown revealed that there were seventeen different functions comprising the production process and that all of the plants were relatively homogeneous with respect to each of these seventeen functions. Processing costs were then obtained for each of the functions in four broad categories: (1) overhead, (2) joint operating, (3) butter manufacturing, and (4) powder manufacturing. These basic data were then divided into three cost elements: (1) capital,

---

[14] This discussion is based upon S. H. Walker, H. J. Preston, and G. T. Nelson, *An Economic Analysis of Butter-Nonfat Dry Milk Plants*, Univ. of Idaho Agr. Exp. Sta., Res. Bul. No. 20 (June, 1953).

[15] Thus the study was not concerned with procurement and distribution (or selling) operations, but only with processing.

(2) labor, (3) and supplies. Each of the seventeen functions and their cost elements were found to contribute to the production process in the same way, except for the differences in the two methods of production—spray and roller-processed powder.

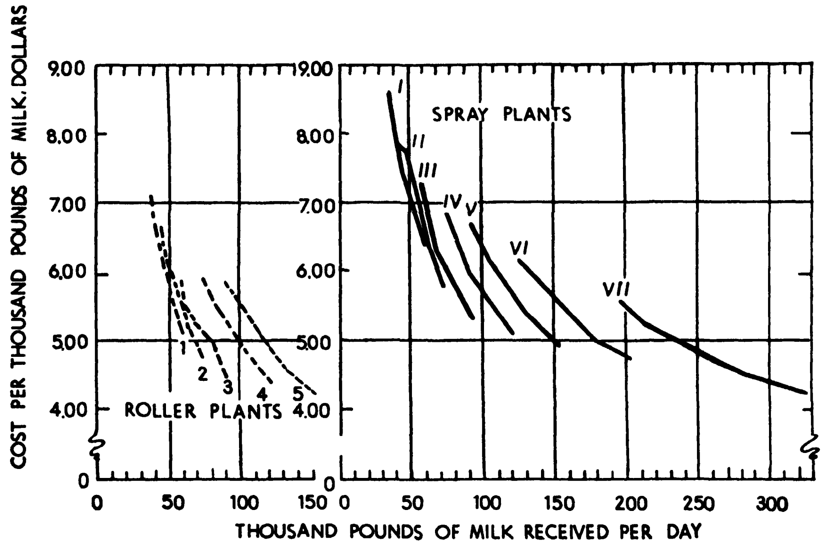
On the basis of the functional cost-volume relationships obtained in this modified type of cross-sectional analysis, detailed plans were developed for the twelve model butter-powder plants, of which five are roller-process models and seven are spray-process models. Seven different sizes of model plants were constructed because the five smallest spray-process models were paired in milk volume with the five roller-process ones for purposes of cost comparison. Although the model plants are simplified and improved versions of the actual plants, they incorporate the principal features of the latter, at least to the extent that most of the functional characteristics are the same in both. In short, the models analyze the fundamental relationships between costs and productive factors to the volume of milk processed by a given plant and to the scale of operations among plants in a multiproduct production process. The models can thus also be used as general standards of efficiency in manufacturing butter and nonfat dry milk.

The Short-Run Curves. Figure 7–8 shows the individual (short-run) plant curves and the relationships between volume of milk and operating costs per 1,000 pounds of milk for each of the twelve model plants in 1948–49. The range in volume per plant is based on seasonal variations in the farm production of milk in Idaho. In each case average costs decline throughout the volume range. Although the rate of decline decreases somewhat as output approaches full capacity, there is no indication of a tendency for average costs to reach a minimum and then rise as output is increased. Perhaps the chief cause of the declining nature of the plant cost curves is the high ratio of fixed to variable costs in butter-powder plants. This appears also to be the principal factor behind the cost curves continuing to decline through all output ranges up to assumed full capacities. With sharply declining average cost curves, marginal costs would be substantially less than average costs throughout the volume ranges of the plants. This constitutes a powerful financial incentive to increase volumes of milk inputs. Examination of the curves also reveals that several plants of different scales can operate at the same level of output. Thus, a larger plant, during a low seasonal flow of milk, processes the same quantity of milk as the smaller plant at the seasonal peak. Three different plants of each type can each process 60,000 pounds of milk per day, but in each of these plants—six in all—there are different average unit costs at this volume. This must mean that the underlying production function of the individual plants differ, and a different combination of productive factors is needed for each of the plants when they are processing this volume. However, each of the plants is operating at a different level of capacity, because of the seasonal variation of milk production at the farm level.

This seasonality of production requires a plant to have more capacity than its average annual volume in order to handle the peak loads of farmers. For example, the average annual output of plants 2 and II is 57,208 pounds, while plants 1 and I can process at full capacity as much as 60,900 pounds per day even though their average due to seasonality is 47,596 pounds per day. The same conditions apply to other plant sizes. The average annual costs of plants 2 and II are $5.56 and $6.84 per 1,000 pounds of milk. Assuming no seasonal variation, these same volumes could have been processed in plants 1 and I at a cost of $5.45 and $6.92 per unit. This means that the seasonality of production costs 11 cents and 22 cents per unit under the conditions depicted in these model plants. These costs are about $2,300 and $4,600 annually, and although they amount to considerable sums, they represent only about $4.00 and $8.00 per year per producer. Hence it is doubtful that the incentive is great enough for producers to smooth out their production through the year.

FIGURE 7–8

Short-Run Cost–Output Functions for Twelve Butter-Powder Plants of Different Capacities

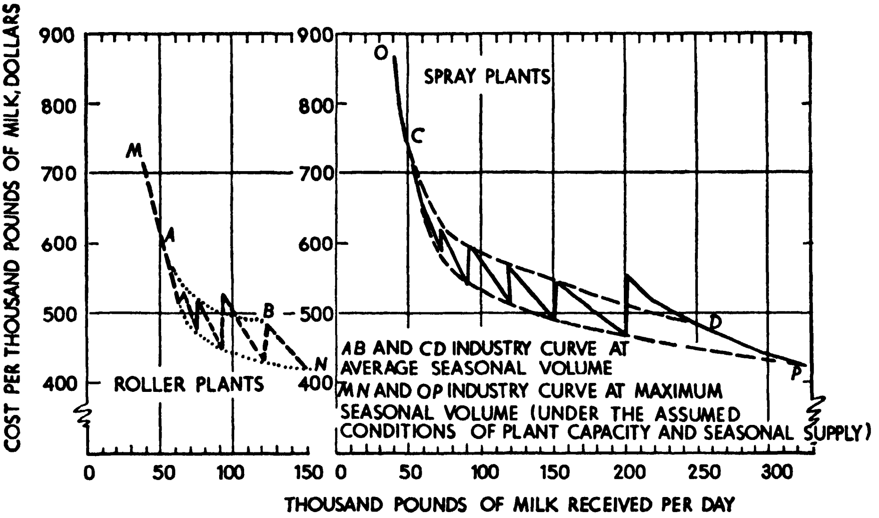


The Long-Run Curves. The short-run curves in Figure 7–8 form a basis for the construction of the long-run planning or envelope curves in Figure 7–9, eliminating the overlapping of the curves. These curves indicate that the industry probably is characterized by long-run decreasing costs, at least in its processing operations. They also help to explain a trend that has long been experienced in the industry—a trend toward fewer but larger plants. With long-run costs declining, there is a pressure upon the industry to eliminate the smallest plants and to construct larger and larger processing units. As for the apparent shape of the industry curve, this is affected by the estimate of the number of plants that can be constructed. In this study seven spray and five roller plants were selected. If it is assumed that these are the only plants that can be constructed, the planning curve will be the irregular line *OP* for the spray process and the line *MN* for the roller process shown in Figure 7–9.

But other combinations of labor and equipment are possible, so this condition is not a rigid one. In this study, the volume is selected by the size of observed churns and by assuming an 8-hour shift in the butter-making function. Further, it is assumed that when two or more churns are employed, they will be of the same size. These assumptions are consistent with observed conditions, but they are not necessary. It is possible to assume that by allowing slight variations in equipment and working periods, more plants of each type can be constructed. With this assumption the smooth line *OP* represents the industry curve for spray process, and the smooth line *MN* for roller process. Since only decreasing costs are depicted, the least-cost point for each plant curve becomes the most important point on the plant curve for constructing the industry curve.

FIGURE 7–9

Long-Run Industry Cost Curves for Butter-Powder Plants



Still another ramification of the least-cost concept, however, is the effect of operating below full capacity for most of the year because of the seasonal pattern of milk receipts. Each plant operates at many levels along the plant curve during each year. For the year, this results in an *average* unit cost higher than the annual average daily volume would indicate if compared directly to the plant curve. (This is mathematically true when the curve opens away from its origin.) A line connecting these points of average annual unit costs for a series of plant curves indicates the average annual long-run industry curve. In Figure 7–9 this is shown by the line *CD* for spray and *AB* for roller process.[16]

Principal Findings. While the detailed findings of the milk-processing study are too numerous to be summarized in a few sentences, the principal conclusions are worth noting from the standpoint of the value of this type of research to management. Also, since the study was concerned both with short-run and long-run cost determination, the following results should be thought of against the general background of advantages to be gained from econometric cost studies of this type, particularly as a basis for future planning by management.

1. Within a given plant, efficiency increases at a decreasing rate, as volumes of milk processed increase up to and including the assumed practical capacity of the plant. (This capacity is less than absolute technical capacity but approximates the concept of capacity generally held in the industry.)
2. Unit costs decline as the scale of operations increases. Under existing technology, butter-powder manufacturing appears to be a decreasing cost industry. The relative decline in unit costs progressively decreases from a high rate among small-scale plants to a lower rate among the larger plants.
3. For the industry, the implications of these conclusions are that, with declining average costs and low marginal costs relative to average costs, there are real economic incentives (pressures) for owners and managers of butter-powder plants to expand plant operations. Similarly, the long-run industry cost curves provide the basis for understanding the long-run trend in the dairy manufacturing industry toward concentration of manufacturing operations into fewer but larger plants.
4. Finally, this study shows in considerable detail (a small portion of which is presented here) the operating costs of butter-powder plants using efficient combinations of resources. The cost and resource standards of the model plants can be applied by managers of existing plants to effect improvements in their operations. It is presumed, of course, that managements can apply the efficiency standards developed in a study of this type to the conditions and facts peculiar to their individual plants.

---

[16] In view of the method of long-run plant selection and planning employed in the study, there may be a question of whether the average annual curves *AB* and *CD* have important differences from the smooth long-run-planning curves *MN* and *OP*. The difference between the two sets of curves must be measured perpendicularly, and when this is done, an important difference appears evident near the output extremes in the spray-process plants. The medium-scale plants exhibit the greatest vertical differences. If this is not the result of chance, it indicates that seasonality affects average costs more adversely in the middle-size plants than in either the small or large plants. Since larger plants were not constructed for roller-type plants, it is impossible to determine whether this same relationship exists in roller-type plants.

Conclusions

The foregoing discussion has come to no definite conclusions regarding the limitations of empirical cost studies. It is appropriate, therefore, to conclude this section with some comments along these lines.

The purpose of deriving statistical cost functions is to isolate, from among the many factors that influence costs, the *net* effect of changes in output rates. In most of the studies that have been made, problems have been encountered that may be classified into two broad categories: statistical and economic. The statistical problems relate to difficulties in methodology and measurement; the economic problems concern the nature and validity of the results. Each may be treated separately for discussion purposes, even though there may be some overlapping in certain respects.

Statistical Problems. The first difficulty relates to the measurement of a diversified output for a firm producing multiple products. Attempts to solve the problem have usually taken the form of weighting the quantity ratios of the various commodities by the relative direct, or variable, costs which they respectively cause. In effect, this amounts to determining output by costs, by introducing a spurious dependence where it is actually the measurement of an independent relationship that is really wanted. Despite this objection, however, it is difficult to see what other solution might be better, for the problem cannot easily be solved regardless of the sort of correlation setup that may be employed. Paralleling this problem is the difficulty of measuring the size of the firm in long-run studies. The more common practice has been to use assets or number of workers, the primary defense being that this provides a convenient way of measuring output by one input. A more accurate measure, however, would appear to be sales, because in sales the various outputs are combined in proportion to their relative importance (prices).

Second are the problems of technological change, and here the difficulties appear insurmountable. Whenever such change occurs, a new cost comes into being, if not a new production function, and no amount of curve fitting will really compensate. At best, the results can only roughly be adjusted as in the Yntema study, rather than accurately accounted for. Closely related to this is the problem of variations in the size of the firm. Assuming that technique remains constant, full harmony exists in the structure of the enterprise when, given marginal costs, minimum average costs cannot be further lowered. This would occur when an increase in fixed costs causes no more than a proportional increase in output rate without increasing average costs.[17] Thus, in the Yntema study, there is the risk of overstating marginal costs to the extent that variations in the size of the Corporation were not taken into account for the observation period. And in the Wylie-Ezekiel study as well, the use of percentage of capacity as the independent variable only *seems* to settle the problem, but not actually so, particularly in the light of the problems of measuring capacity even in ordinary plants.[18] What most analysts have done to solve this problem has been to avoid it, although carefully, by choosing firms and periods in which technological change and variations in size are absent. Thus, Dean, exclusive of the excellent nature of his statistical work, states in his leather belt shop study: "The period . . . was chosen because if fulfilled the following conditions most satisfactorily: (1) The rate of output and other measurable determinants varied sufficiently to yield observations over a wide range . . . (3) The plant and equipment remained unchanged during the analysis period, permitting the observation of short-run adjustments uninfluenced by long-run changes. . . ."[19] The same criticism could be leveled against most other authors, but there is no need to belabor the point. The conclusion to be drawn is that when a problem is solved by avoiding its inherent difficulties, the solution is usually not a very satisfying one.

---

[17] This has been called Schneider's "law of harmony," discussed in a well-known empirical study of costs for a German cement firm, made by Ehrke and Schneider. See K. Ehrke, *Die Uebererzeugung in der Zementindustrie von 1858–1913*

A third statistical difficulty lies in the measurement of costs. Since the general problems have been discussed previously in earlier chapters, the more common ones may be stated briefly. With respect to asset valuation, there is the problem of valuing raw materials (inventory) at cost or market. One argument is that they should not be valued in a cost study because they are not part of production. The other argument is that they should be valued, because an index of successful management is its ability to buy raw materials at low prices. A further problem is the valuing of land. Accountants would value it at its definite and objective historical cost, but economists would hold that historical costs are irrelevant for decisions affecting the future. Nor would economists hold to current market value, because this is approximately equal to discounted future earnings, and the firm would always be earning the going rate of return on investment. (See also Chapters 10 and 11.)

Fourth, there are statistical considerations to be made in choosing a measure of efficiency. For example, if rate of return on investment, a common measure, is used, companies paying high executive salaries instead of high dividends in smaller owner-officer corporations will appear inefficient. One approach is to accept the corporation's decision as final; another

---

("Overproduction in the Cement Industry, 1858–1913"), Jena, 1933, p. 290. The first formulation of the law was in an article by Ivar Jantzen for which a German translation appears in E. Schneider, *Theorie der Produktion* ("Theory of Production"), Vienna, 1934, pp. 83–92.

[18] See George Terborgh's enlightening remarks in, "The Problem of Manufacturing Capacity," *Federal Reserve Bulletin* (July, 1940), pp. 639–46.

[19] Page 11; see the bibliographical note at the end of this chapter.

is to adjust the salaries of officer-owned corporations to equality with non-officer-owned corporations of equal size. Both methods are questionable, however, and the results of the two quite different.[20]

Still other problem areas could be mentioned, but enough has already been said to indicate the sort of difficulties frequently encountered in the empirical study of costs. It should be evident, therefore, that although studies of this kind can provide a useful guide for management planning, ambiguities are attached to almost all econometric studies and their practical value will depend upon how carefully they are interpreted.

Economic Problems. Most of the analysts who have investigated statistical cost functions, as in the Yntema study, have found a tendency for short-run total costs to be linear and marginal costs to be constant. Since this seems to contradict certain assumptions of economic theory, could it be that the theory is incorrect, unrealistic as to the facts, and hence in need of basic revision? Various closely related explanations can be offered, all of which appear to contribute to the correct answer.

1. The assumptions of economic theory are approximately correct, but total costs will also be linear or nearly so within this range. In empirical studies, wider ranges could be covered closer to the output extremes of the total cost curve, the curve would bend in the end areas and thus yield decreasing and increasing marginal costs at these extremes.

2. The assumptions of economic theory are approximately correct, in that constant marginal costs prevail in industry, at least over wide ranges of total cost. If this is true, it means that within the relevant range of the data there is a constancy of factor proportions and therefore no significant economies or diseconomies of large-scale production. This leads to the inference that, in the final analysis, the only comprehensive test of efficiency is survival. If small firms tend to disappear and large ones survive, as in the automobile industry, we must conclude that small firms are relatively inefficient. If small firms survive and large ones tend to disappear, as in the textile industry, then large firms are relatively inefficient. In reality, however, we find that, in most industries, firms of very different sizes tend to survive, and hence we conclude that usually there is no significant advantage or disadvantage to size over a very wide range of outputs. In other words, it seems plausible to conclude that, in many different industries, constant returns to scale are a good approximation to reality.

3. The cost curves of economic theory are static and hence can at best provide only an approximate explanation of the organization of enterprise in a fluctuating dynamic economy. When firms have to contend with business cycles, they must of necessity be flexible so that they can adapt to changing business conditions. This means that they must be able to produce efficiently over a wide (normal) output range, and this in turn

---

[20] See J. L. McConnell, "Corporate Earnings by Size of Firm," *Survey of Current Business* (May 1945). See also G. Stigler, *The Theory of Price*, rev. ed., pp. 143–44, for some further theoretical implications concerning empirical studies.

requires flat or nearly flat average and marginal costs at least within that range.[21]

ADVERTISING COSTS—BUDGETING

In Chapter 5 the problem of discovering the nature of the sales-advertising function was discussed, primarily from the standpoint of controlling and manipulating demand. In this section we turn our attention to advertising again, but this time with emphasis on the economic implications of planning and allocating advertising costs both under short-run and long-run conditions. Although an analysis of advertising costs is frequently treated as part of the study of distribution costs, the two are separated here in order to facilitate discussion. Some comments on distribution costs are reserved for the following section of this chapter.

It will be recalled from Chapter 5 that a distinction exists in economic theory between production costs and sales costs. Production costs are those resulting from the production of the product itself; sales costs arise from those activities designed to influence the demand curve, as when firms incur expenses of salesmen, public relations, gifts to purchasing agents, and the use of various advertising media such as radio, television, newspapers, and so forth. In theory the distinction between the two classes of cost is usually clear cut; in practice, a sharp line between the two cannot always be drawn because some costs, e.g., packaging, may fall partly in each category. Nevertheless, the distinction is conceptually useful, at least as a beginning to the discussion of short- and long-run budgeting, even though some expenditures will occasionally defy precise classification.

Short-Run Budgeting

Three approaches to the short-run budgeting of sales costs may be outlined: (1) the incremental method, (2) per-cent-of-sales method, (3) objective-and-task method. The first draws entirely on principles of economic theory and is useful as a guide to thinking about the subject as well as pointing out clearly where empirical research is really needed. The remaining two are the methods in common use by advertisers today.

Incremental Method. An important assumption underlying the incremental method (or the marginal method in its most refined form) is that the firm will seek to adjust its selling expenditures to the level which will allow maximum profit. Various methods for attacking the problem have been developed by a number of economists, notably Boulding, Buchanan, and Chamberlin.[22] In all instances three variables are involved: price, selling costs, and sales. Sometimes either price or selling cost is varied with the other held constant to determine the effect on sales; sometimes both price and selling cost are varied and the effect on sales is noted. In all cases the ultimate objective is to arrive at a combination of price and selling costs that will result in a sales figure which yields maximum net profit to the firm.

---

[21] See Tintner, *Econometrics*, pp. 47–49; O. Lange, *Price Flexibility and Employment*, pp. 2 ff.; and L. Robbins, *The Great Depression*, for further observations of this from various standpoints.

It is not within the scope of this book to go into the details of the theoretical structure concerning the nature of these adjustments. However, some essential theoretical concepts can be presented based on the principles of production economics presented previously in Chapter 6.

Basically, the procedure used is to regard selling cost as a kind of productive resource, variations in which will cause changes in sales if price is held constant. There are thus two classes of problems: first, to determine the optimum total selling cost expenditure to sales; second, given the total selling cost expenditure or budget, to determine the optimum allocation of that expenditure among competing advertising media. The first problem is akin to the factor-product type of analysis in Chapter 6 where a simple relations type of production function was constructed, except that now the factor would be varying doses of selling costs for a homogeneous medium (instead of labor as in the auto laundry study) and the product would be sales (instead of cars washed per hour). The result should take the form of a selling cost–sales function somewhat similar to the production function of Figure 7–1, although it may of course be linear over a range of inputs. Similarly, if both price and selling costs are varied, the principles parallel the factor-factor type of analysis where the object is to determine either (1) the optimum combination, of price and selling cost to produce a given sales level, or (2) the maximum sales level that can result from a given combination of price and selling cost.

The second problem, that of allocating a given budget among competing media, is also an extension of principles derived from production economics. Using the incremental notation Δ, let ΔS_1, ΔS_2, ΔS_3 . . . ΔS_n represent the increase in sales at a given price from advertising media 1, 2, 3 . . . n, and let ΔA_1, ΔA_2, ΔA_3 . . . ΔA_n denote the additional expenditure sum on various forms of advertising, say in $1,000 blocks. For (equilibrium) optimum allocation of expenditures, it is necessary that the corresponding incremental ratios be equal. That is, if medium 1 represents radio time, medium 2 is spot television commercials, etc., then optimum allocation requires that

$$\frac{\Delta S_1}{\Delta A_1} = \frac{\Delta S_2}{\Delta A_2} = \frac{\Delta S_3}{\Delta A_3} \cdots \frac{\Delta S_n}{\Delta A_n}.$$

---

[22] Their works are well known to professional economists, particularly the successive editions of K. Boulding's *Economic Analysis*, E. Chamberlin's *Theory of Monopolistic Competition*, and N. Buchanan's "Advertising Expenditures, A Suggested Treatment," *Journal of Political Economy* (August, 1942), pp. 537–57.

For if the ratios are not equated, as when

$$\frac{\Delta S_1}{\Delta A_1} > \frac{\Delta S_2}{\Delta A_2},$$

medium 1 in this case would be preferred to medium 2. Hence it would pay to reduce (or withdraw) expenditures on medium 2, thereby raising the value of that ratio, and increase expenditures on medium 1, thereby reducing the value of that ratio. When the ratios are equal, net profit is at a maximum.

Theory thus leads to the conclusion that basically similar principles are involved in regarding sales costs in the same manner as the hiring of productive factors. Further, it infers that there is a law of diminishing returns to advertising; in that constant increments of advertising expenditure shift the demand curve to the right in ever-decreasing amounts (assuming price is unvaried) so that eventually the elasticity of advertising approaches zero.

These conclusions, plus the advantages cited above—that the incremental method provides a guide for thought clearing and empirical measurement—are offset by some limitations: (1) The theory makes no allowances for the investment or cumulative effects of advertising, nor does it recognize the significance of lagged responses; (2) it takes no account of the effect of competitors' reactions to advertising; and (3) it assumes that the effect of advertising on sales volume can be measured so that the results can serve as a basis for budgeting. Concerning this last limitation, it may be noted that some empirical studies have yielded good results, particularly with respect to the allocation problem. Using the incremental concept outlined above, controlled experiments employing Latin-square designs of the type discussed in Chapter 3 were set up for various test markets. In one such study, five different media treatments were taken and the treatments were rotated between the relevant cities during the week. Covariance analyses were then conducted and tests of significance were made. The results of the study revealed significant differences in advertising media sufficient to warrant a reshuffling of the company's selling costs for its next fiscal period, with substantially larger sales and profits as a result. But in this case the firm was particularly suited to this type of analysis because its product was a service—a unique type of insurance plan for television sets—on which it had a near-monopoly status in its own regional area. Normally, however, the incremental method could not as easily be adapted to other business firms because of the limitations cited above, although the use of mail order and of keyed responses techniques offers another area in which the approach could be successfully developed. In view of these limited applications, most companies normally employ either of the other two approaches to short-run budgeting discussed next.

Per-Cent-of-Sales Method. The per-cent-of-sales method is self-explanatory. It consists of taking a fixed percentage of the previous period's sales or of an average of several periods' sales and using this as a budget for the company's next fiscal period. An alternative approach is to select a percentage based on a forecast of sales. In either case, the shortcoming of the method is essentially that it places the cart before the horse, by not recognizing that advertising expenditures are made for the purpose of influencing sales. Using a percentage figure based on past experience gives no indication of how much should be budgeted to *increase* future sales, i.e., shift the demand curve to the right, which is the way in which advertising should be viewed. Used in its present manner, the per-cent-of-sales approach to advertising budgeting is more an effect of sales rather than a cause. Its widespread use, however, is probably due to the fact that it offers a simple and mechanical method of budgeting and control, and that it permits the advertising expense to pay its own way because the expenditure fluctuates according to sales.

Objective-and-Task Method. In this method, the sales objective is established first, usually on the basis of the difference between the sales level that would be expected with and without advertising. This difference becomes the objective, and the "task" is to determine the advertising budget needed to reach the objective. In this simple form, the method suffers from the weakness that it takes no account of whether the predetermined increase in sales—the objective—is worth the increased cost needed to attain it, i.e., whether or not the ultimate result will be an increase in net profit. If the appropriate measures could be obtained for comparison and evaluation, the objective-and-task method would come closer to the incremental approach described above. Unfortunately, such measures are usually difficult to establish, and the result is that the approach is used by most companies in its simpler form. In fact, about three quarters of American advertisers were using this method in the postwar period.[23]

Long-Run Budgeting

The above discussion of short-run budgeting took no account of business cycles and the bearing they may have on a firm's advertising expenditures over the years. It is a truism, of course, that a company's long-run advertising outlay will either be constant or it will fluctuate. If it fluctuates, there are a wide range of possibilities, but the more likely ones would probably be that the rate of change is either greater than, equal to, less than, or in some manner inversely proportional to, the firm's sales or profits. The problem, then, is to see whether a set of principles can be established that would serve as a guide to management in formulating a cyclical advertising policy.

---

[23] See *Printers' Ink* (December 28, 1946), p. 26. Also, a much fuller analysis of the approach is given by A. Haase, *The Advertising Appropriation.*

Unfortunately, no systematic program can be presented that would be applicable to all firms, partly because of the differences among companies as to size, products, and marketing methods, and partly because economists have still much to learn about the nature and causes of business cycles. Regarding the latter, many single-cause theories have been proposed, but economists generally agree that cycles are the result of a multiplicity of causes rather than any single factor, which places a chief obstacle in the way of developing a unified budgeting policy.[24] In view of this, we can only point out a few of the important concepts gained from business cycle theory and experience that may be useful as a tool to management when considering long-term advertising policy. The chief task is to see if certain guides are available that can be used to improve the efficiency of the advertising expenditure over the business cycle. Some key guides may be listed as follows:

1. The income elasticity of demand, which measures the percentage change in quantity demanded resulting from a given percentage change in income, when applied to particular products or classes of products (see Chapter 5), might serve as a useful indicator. Products with a high elasticity, such as durables and luxuries, would require heavier advertising to overcome consumer resistance when incomes are low. Other elasticity measures discussed in Chapter 5, such as price, substitute, and promotional elasticities, and their interrelations, can provide further quantitative evidence on which to base a cyclical advertising strategy.

2. A rational program for timing product improvement and new-product development, combined with the appropriate type of advertising depending on the phase of the cycle, offers a further area for improved sales-cost budgeting. The current considerations on the part of some aumobile manufacturers to turn out smaller and more economical cars, in anticipation of consumers becoming more cost conscious, may turn out to be a case in point.

3. A planned program of "investment advertising" over the full course of the business cycle can yield two broad advantages. (*a*) In depression periods it can help maintain consumers' brand preferences at a time when price competition is relatively more severe. This assumes that it would cost more for the firm to regain lost buyers through advertising than it would to maintain at least a minimum level of advertising in de-

---

[24] An example of one attempt, based on the experience of one company, is O. Keyser, "A Counter-Cyclical Fund for Advertising," *Advertising and Selling* (April, 1947). Keyser suggests that a fund be accumulated in prosperity periods from which constant advertising expenditures could be maintained over the cycle. He assumes a "psychological theory" of business cycles, i.e., that optimistic and pessimistic errors, once under way, are self-generated in an endless chain, as expounded in the 'twenties by such economists as A. C. Pigou, *Industrial Fluctuations*, and F. Lavington, *The Trade Cycle*. See also, however, R. S. Vaile, "Use of Advertising During Depression," *Harvard Business Review*, Vol. 5, and L. C. Wagner, "Advertising and the Business Cycle," *Journal of Marketing* (October, 1941), for further viewpoints.

pression periods. (*b*) In prosperity periods, the firm that has strongly entrenched itself by continuous advertising can exploit the fact that buyers are less price conscious, while simultaneously incurring some savings in sale costs by not having to match the heavier advertising expenditures of competitors at a time when media charges (e.g., newspaper space rates) are more expensive.

In summary, the proposals amount to the suggestions that: (1) firms direct their long-run advertising strategy to exploiting the various elasticity characteristics of particular products and product classes; (2) firms time their rate of product development and improvement to accord with the need for effective advertising in depression periods; and (3) firms stabilize their selling expenditures somewhat by cutting off the peaks and troughs so as to maintain continuous advertising over the business cycle.[25]

A NOTE ON DISTRIBUTION COSTS

The study of distribution costs from the firm's-eye viewpoint has become an increasing concern of business economists and marketing analysts since the postwar years. Perhaps a chief reason has been the development of operations research procedures, particularly mathematical programming techniques, that are especially well suited to the solution of problems involving distribution costs. In this section a few brief comments will be made concerning the techniques that management can profitably use in the area of distribution cost analysis.

The Problem

The main objective of distribution cost analysis is to reduce distribution costs that are otherwise out of line because of a misallocation or maldistribution of marketing effort. Ultimately, the hope is that enough information can be obtained, particularly with respect to cost data, on the basis of which the firm's marketing expenditures and resources can be reallocated so as to yield maximum profits. The problem exists primarily because, in most businesses, whether manufacturing, retailing, or wholesaling, there is a distorted pattern of relationship between the costs and profits attached to each segment of the business, to items in the product line, to customers, orders, and territories, and to selling, advertising, and other marketing efforts. The result is that for many if not most firms, a great majority of the customers may be responsible for a very small percentage of sales volume, or a large per cent of products manufactured may account for hardly more than a few percentage points of total sales. Evidently, management fails to recognize that each dollar's worth of marketing effort, in terms of salesmen's time, advertising, warehouse space, etc., should be directed to where it yields the largest increment in net profit. This failure causes a disproportionate spreading of marketing effort, which in turn results in the company's profits as a whole being substantially less than they might otherwise be if marketing resources were reallocated in a more efficient manner.

---

[25] For an example of the significance of timing product development, see "How Philco Doubled Sales During the Depression," *Printers' Ink* (October 22, 1937), p. 17. On the significance of continuous advertising, see "Why Advertising Should Be Continuous," *ibid.* (April 28, 1938), p. 12, which contains a number of policy statements on the matter by leading corporation executives.

The Solution

In theory, the approach to the solution of the problem could be framed in terms of an extension of the incremental ratio notion similar to the procedure discussed in the previous section dealing with the allocation of advertising effort. This will be done below. In practice, however, it is difficult to determine which parts of the firm's marketing process contribute to its costs, sales, and profits. Unfortunately, prevalent accounting techniques do not provide satisfactory answers, and although most businessmen think they do, the fact is that they are laboring under a serious misapprehension. The typical accounting procedures for recording the results of marketing activities are not sufficiently detailed because they show only the averages; further, their information is distorted by arbitrary cost calculations and their figures are only part of what is actually required. The correct approach to the solution of the problem, therefore, lies first in providing a finer breakdown and a reclassification of the company's average cost and profit data. The firm's over-all distribution costs must be allocated to the *specific segments* of the business *for which they are incurred.* Thus, the sale of 100 dozen watches to medium-sized retail jewelers in a particular city may require x dollars worth of salesman time, y dollars in transportation and warehousing costs, z dollars in advertising expenditure, and so on. By segmenting the cost and profit data, the net profits or losses for each segment can be calculated separately. The object, therefore, is to divide the business of the company into segments classified, for instance, by categories of customers and products, and then to determine the marketing costs, production costs, and net profits or losses for each segment separately. There are thus two basic principles and techniques that may be summarized:

1. The distribution expenses of the firm should be reclassified from a *natural* expense basis into *functional* cost groups, bringing together all of the indirect costs associated with each marketing activity or function performed by the company.
2. The functional cost groups should be allocated to products, customers, and other segments of sales according to measurable factors, or product and customer characteristics which bear a cause-effect relationship to the total amounts of these functional costs.[26]

---

[26] See C. H. Sevin, "Cost Control in Selling by Manufacturers," *Cost and Profit Outlook* (May, 1957); also, W. J. Baumol and Sevin, "Marketing Costs and Mathematical Programming," *Harvard Business Review* (September–October, 1957).

In order to perform the required calculations for each segment and each functional cost group, it is necessary to distinguish between three classes of *marketing* costs for which data are needed.

Common Fixed Distribution Costs. These are costs that are incurred in common for different sales segments, and their magnitudes do not vary with the volume of sales in any one segment. An example is the advertising expense of a company's name, which probably influences sales in all segments in varying degrees. These costs are excluded from the distributional cost and profit analysis.

Variable Distribution Costs. These are distribution costs that vary with sales and hence can be allocated. An example is the increased freight bill resulting from additional sales, which is clearly variable and can be readily assigned. Some variable and fixed marketing costs are more difficult to distinguish, however. Warehousing cost, for example, is a fixed cost when not used to capacity, but becomes variable when storage space fills up and management considers the construction of more space to eliminate a bottleneck. Despite the difficulties of segregating costs, it is necessary since the variable costs must be included in the analysis.

Separable Fixed Distribution Costs. These are fixed marketing costs that can and should be allocated to sales. For example, the value of a sales manager's time devoted to a particular sales segment is variable, although his salary is not. Hence, his time should be allocated even though his salary is a fixed cost. It follows that the incremental cost of separable fixed expenses, such as the cost of the manager's time required to make *additional* sales in each segment of the business costed, should be computed if possible and these figures should be kept distinct from variable costs.

When these figures are obtained, marketing effort can be redistributed to yield greater profits. In words, the condition for maximum profit is that marketing effort be reallocated to those segments of sales where an additional unit of marketing effort will yield the largest contribution to net profits and overhead, after deduction of variable costs. In symbols, letting ΔP represent incremental contribution profit defined as the difference between incremental sales, ΔS, and incremental variable costs, ΔVC, and letting ΔR denote the additional resource or effort devoted to a sales sector, effort should be increased in a sector until

$$\frac{\Delta S - \Delta VC}{\Delta R} = \frac{\Delta P}{\Delta R} = \text{maximum}.$$

Using the subscripts 1, 2, 3 . . . *n* to code each market sector, such as New York, Chicago, San Francisco, etc., optimum allocation of marketing effort between sectors requires that

$$\frac{\Delta P_1}{\Delta R_1} = \frac{\Delta P_2}{\Delta R_2} = \frac{\Delta P_3}{\Delta R_3} \cdots \frac{\Delta P_n}{\Delta R_n}.$$

For if the ratios are not equal, as for example if

$$\frac{\Delta P_1}{\Delta R_1} > \frac{\Delta P_2}{\Delta R_2},$$

it implies, as in all types of resource allocation problems, that net profits can be increased by reducing or withdrawing effort in market sector 2, thereby raising that ratio, and increasing it in market sector 1, thereby reducing that ratio.

The analysis thus brings to the forefront the notion that there is a law of diminishing returns with respect to the allocation of marketing effort—an assumption that is not unreasonable by any standards and certainly in accord with the experience of business firms. Framed in this manner, the analysis is capable of a practical solution by employing methods of mathematical programming. But this is part of the science of operations research, a full discussion of which is beyond the scope of this book.

BIBLIOGRAPHICAL NOTE

The classic work on the study of costs and certainly one of the most important contributions both to economics and business administration in the twentieth century is J. M. Clark's, *Studies in the Economics of Overhead Costs.* Chapter III contains a classification of cost concepts more extensive than the one presented here. On cost measurement, in addition to the works mentioned in the footnotes, there are the well-known studies by Dean, *Statistical Cost Functions of a Hosiery Mill; The Relation of Cost to Output for a Leather Belt Shop;* and *Statistical Determination of Costs, with Special Reference to Marginal Costs.* An interesting illustration of the engineering approach is a cost function derived for an airline based on principles of aeronautical engineering, by A. Ferguson, "Empirical Determination of a Multidimensional Marginal Cost Function," *Econometrica* (July, 1950). Some critical evaluations of econometric cost analyses are given by C. R. Noyes, "Certain Problems in the Empirical Study of Costs," *American Economic Review* (1941); and by H. Staehle, "The Measurement of Statistical Cost Functions: An Appraisal of Some Recent Contributions," *ibid.* (1942). Further comments concerning both theory and measurement appear in: J. Mosak, "Some Theoretical Implications of the Statistical Analysis of Demand and Cost Functions for Steel," *Journal of the American Statistical Association,* Vol. 36; in a note by G. Stigler, *American Economic Review* (March, 1940); and in the latter's "Production and Distribution in the Short Run," *Journal of Political Economy* (June, 1939). Some brief comments are also made in Tintner, *Econometrics,* chap. 2. Finally, on sales costs, in addition to the path-breaking work of Chamberlin, especially chaps. V–VII, and the works mentioned in the footnotes, there are: F. P. Bishop, *The Economics of Advertising;* F. A. Lever, *Advertising and Economic Theory;* J. P. Hayes, "A Note on Selling Costs and the Equilibrium of the Firm," *Review of Economic Studies* (1944–45); and H. Smith, "Advertising Costs and Equilibrium," *ibid.* (1934).

For more concise and perhaps less technical reading, there is a classification of cost concepts in Dean, chap. 5, in an earlier work by C. T. Devine, *Cost Accounting and Analysis*, chap. XXX, and in Howard, chap. VII, none of which are as complete, however, as Clark's pioneering work mentioned above. Colberg, Bradford, and Alt, chaps. 4 and 5, in addition to the above texts, provides supplementary reading as well. On advertising, a standard general work is N. Borden, *The Economic Effects of Advertising*, while Dean, chap. 6, and Howard, chap. XIII, present more succinct treatments from the management standpoint.

QUESTIONS

1. What is the use of distinguishing between various cost concepts, since most of them are not directly obtainable from the company's accounting records?
2. Distinguish between absolute or outlay costs and alternative or opportunity costs.
3. What is the difference between direct and indirect costs?
4. (*a*) Define, in your own words, the terms "fixed cost" and "variable cost," and give examples. (*b*) Do fixed costs ever "vary"? Are variable costs ever "fixed"? Explain.
5. Express the relationship between fixed cost, variable cost, total cost, and marginal cost.
6. (*a*) Define short-run costs and long-run costs. (*b*) What is the practical usefulness of distinguishing between fixed and variable costs, and between short- and long-run costs?
7. Explain the nature of the distinction between differential or incremental costs and residual costs.
8. State briefly the basic differences between sunk, shutdown, and abandonment costs, and give examples.
9. Distinguish between urgent and postponable costs and between escapable and inescapable costs.
10. What is the nature of the distinction between replacement and original cost?
11. "There is no such thing as a cost of production." Discuss.
12. (*a*) What determines the actual shape or curvature of a firm's cost curves? (*b*) What is the effect of a change in factor prices on a firm's cost-curve structure? (*c*) Explain why management might be interested in having empirical estimates of its short- and long-run cost-output curves.
13. (*a*) How are the short-run cost curves derived? (*b*) The long-run average cost curve? Explain.
14. "For outputs less than the long-run optimum level, it is more economical to 'underuse' a slightly larger plant operating at less than its minimum-cost output." Illustrate this proposition graphically, and also its converse.
15. Outline the nature of the preliminary considerations involved in establishing the analytical framework for cost measurement.
16. What factors are involved, in addition to those brought up in question 15, in measuring long-run costs?

17. (*a*) What use can management make of the type of results obtained in the cost studies discussed in this chapter? (*b*) Outline the nature of the problems typically encountered in making empirical cost studies.
18. (*a*) Distinguish between the common approaches to advertising budgeting employed by business firms, and evaluate each. (*b*) "It pays to increase advertising as long as this results in increased sales." Discuss. Rephrase if necessary.
19. Outline the type of proposals that may be suggested as a basis for establishing a long-run advertising budgeting program.
20. (*a*) What are distribution costs? (*b*) Explain the nature of the problem of distribution costs from the management standpoint. (*c*) What is the advantage of allocating distribution costs into functional cost groups?
21. Donald R. G. Cowan, in the *Michigan Business Review* (May, 1958), points out the following eleven approaches to the analysis of distribution costs. Explain briefly the type of research you would expect to be done in adopting each of these approaches. (*a*) Product approach. (*b*) Product-line approach. (*c*) Distribution channels approach. (*d*) Engineering approach. (*e*) Operations research approach. (*f*) Accounting approach. (*g*) Economic approach. (*h*) Personnel approach. (*i*) Organizational approach. (*j*) Standardization approach. (*k*) Management approach.

Chapter 8

PRICING: PRACTICES AND POLICIES

Having completed an analysis of demand, production, and costs, and their role in forward planning by management, we turn our attention now to the area of pricing. The purpose throughout, as in previous chapters, is to develop concepts that are conducive to measurement and that will serve as a guide for more effective control by executives. For the most part, attention will be directed to the pricing practices and policies of manufacturing firms in oligopolistic markets, but some space will also be devoted to other problem areas that have gained in relative significance in recent years. Selected topics will be considered that seem to be of greatest relevance to the practical pricing decisions of top management in the light of current thinking on the subject.

PRICING CONCEPTS AND MARKETING POLICIES

The theory of monopolistic competition is well over twenty years old, yet relatively little of it has worked its way into the standard textbooks of business administration in general and of marketing in particular. Hence it has also gained little headway in the thinking of businessmen, especially with respect to pricing. Various price concepts such as "odd prices," "customary prices," "price lining," etc., are discussed in elementary marketing texts, but purely from a descriptive standpoint without any indication of the relation of these notions to economic theory. It is appropriate, therefore, to begin this chapter on pricing with a discussion of various marketing price policies, paying particular attention to illustrating the fact that these policies are merely special cases of the general theory of monopolistic competition. The chief advantages of presenting these concepts in an analytical framework are that the theory provides (1) a sounder basis for discussion and evaluation, and (2) an indication of what it is that really needs to be measured. The procedure followed will be to present, first, a few statements about the general theory of pricing that are now common to elementary economics textbooks but which stem originally from the writings of Chamberlin and Robinson in the early 'thirties; and second, an illustration of various common pricing concepts from the viewpoint that they represent nothing more than an implicit assumption regarding the nature of the particular demand curve.

General Theory of Pricing

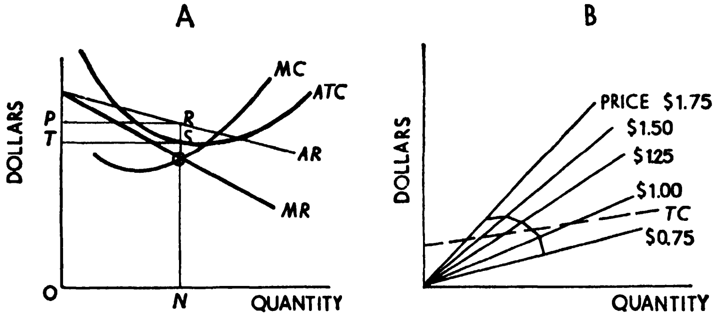
The general theory of pricing, i.e., the theory of pricing under monopolistic competition, is basic to all the pricing concepts, practices, and policies discussed not only in this section but throughout most of this chapter. The fundamental scheme, therefore, may be sketched briefly as follows.

1. Figure 8–1A represents the demand and cost structure of a firm under monopolistic competition. Because of product differentiation, the seller will have some degree of monopoly power. This is indicated by the negatively sloping demand or average revenue curve, *AR*, showing that at higher prices, the firm loses some sales but not all sales (as compared to a seller in pure competition for whom the demand curve is horizontal). From this the marginal revenue curve, *MR*, representing the change in total revenue resulting from a unit change in output, is derived. Similarly, total cost per unit of output, or average total cost, *ATC*, and the marginal cost curve, *MC*, are shown. The condition for establishing the output that will yield maximum short-run profit is then determined: the firm should produce to where marginal cost equals marginal revenue, which is where the rate of change in total costs equals the rate of change in total receipts. In terms of the diagram, the most profitable output is *ON* units, to be sold at a price of *OP*(=*NR*) dollars per unit, which, as determined by the demand curve, is the highest price per unit that can be charged to clear the volume *ON*. The total receipts would then be the price times quantity or the area of the rectangle *OPRN;* the total cost is the average cost times the number of units or the area *OTSN*. Net profit is thus the area of the rectangle *PRST*, or the difference between the larger (total revenue) and smaller (total cost) rectangle. This net profit rectangle is

FIGURE 8–1

GENERAL THEORY OF PRICING

A, Traditional Theoretical Formulation; *B*, Break-Even Formulation



maximized only when production is carried to the $MC = MR$ level. At any other output, the area of the rectangle would be smaller under the given demand and cost conditions.

2. Businessmen, accountants, and engineers, who are more familiar with break-even charts, will prefer a translation of these theoretical principles into the break-even formulation of Figure 8–1B. Some modifications of the chart are necessary, however, because the conventional break-even chart shows only the various sales possibilities at a *single price* and the volume required to break even at that price (see Chapter 4). In Figure 8–1B, several total revenue curves (TR) are drawn, each assuming a different price per unit. Each TR curve thus reveals the total receipts that would be realized over a range of sales at a given price. The lower the price, the flatter the TR curve, indicating a wider sales range; similarly, higher prices indicate steeper TR curves and narrower sales ranges, evidencing some degree of monopoly power on the part of the firm due to product differentiation or other factors. On each TR curve a point is estimated showing the sales volume actually realized at that price as a result, say, of a controlled price experiment in several markets. The locus of these points is then the curve DD', which may be thought of as a kind of demand curve except that it shows total revenue rather than average revenue as does the usual demand curve such as AR in Figure 8–1A. The condition for profit maximization would then be to maximize the vertical distance between DD' and the total cost curve TC. This formulation thus yields an actual advantage over the traditional marginal cost–marginal revenue formulation of Figure 8–1A because it shows not only the correct price and volume, but also total cost, total revenue, and net profit. In the following discussion of price policies, however, the conventional AR curve is used because it conveys the various concepts more clearly and precisely.[1]

Odd Prices

It is a practice among some companies, particularly noticeable in retailing, to set prices in such a way that they end either in an odd number or just under a round number. The assumption is that it is possible to sell a greater number of items priced, for example, at 23 cents rather than 20 cents, or at 98 cents rather than 89 cents. The idea applies to higher-priced merchandise as well, such as a suit of clothes at $79 instead of $75, or an automobile at $1,995 rather than $2,000. These notions, expressed in virtually all marketing textbooks, can be illustrated more meaningfully by the use of the average revenue or demand curve of economic theory.

---

[1] The theoretical formulation of Figure 8–1A could also have been presented in terms of the total revenue and total cost curves of economic theory, but the method shown is more common, and often more practical. (See the reference to E. R. Hawkins in the bibliographical note at the end of this chapter, concerning some of the discussion in this section.)

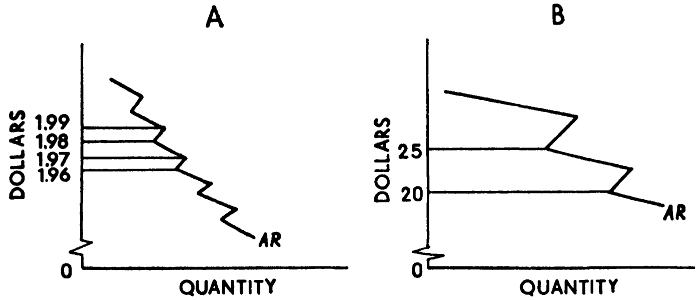
Figure 8–2A shows the type of demand curve *assumed* by a seller who sets his prices by the odd number "rule," i.e., that sales are larger when the price ends in an odd number than when it ends in the next lower even number. Thus the quantity demanded is greater at $1.99 than at $1.98, and greater at $1.97 than at $1.96. Although this type of demand curve must be believed to exist by sellers who price by this odd-number method, there is no conclusive evidence that an odd-price policy actually results in larger sales as is assumed by those who employ it.[2]

The second concept, that of round-number pricing, is illustrated in Figure 8–2B. The assumption here is that sales will be larger when the price is set just under a round number or critical point, such as $20 or $25 in the diagram. At prices slightly below these critical points, demand is elastic in that a small decrease in price from the critical point brings a more than proportional increase in sales, regardless of whether the price ends in an odd or even number. As with odd-number pricing, this assumption about the shape of the demand curve by those who practice round-number pricing has never been subjected to any extensive tests.

FIGURE 8–2

ODD PRICES

A, Odd-Number Pricing; *B*, Round-Number Pricing



Psychological Prices

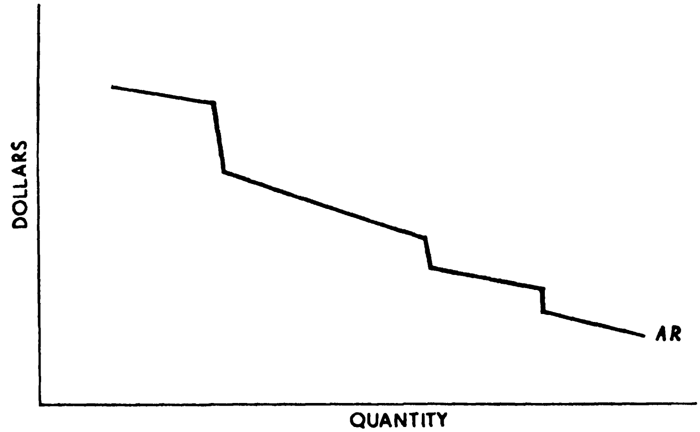
Somewhat similar to the demand pattern in Figure 8–2 is that found in Figure 8–3. The phenomenon represents the type of demand curve that would exist for what some writers have called "psychological prices." It has been found in some pricing experiments that a change in price has little effect over certain ranges of demand, thus yielding a step-type of *AR* curve. This differs from the concept of odd pricing in that the curve need not have any positively inclined segments, and the critical points are not necessarily located at each round number but only at prices that are psychologically significant to buyers. Some pricing experiments at the R. H. Macy department store in New York have revealed such step-

---

[2] For the most extensive study made of the subject, see E. Ginsberg, "Customary Prices," *American Economic Review* (1936).

FIGURE 8–3

Psychological Prices



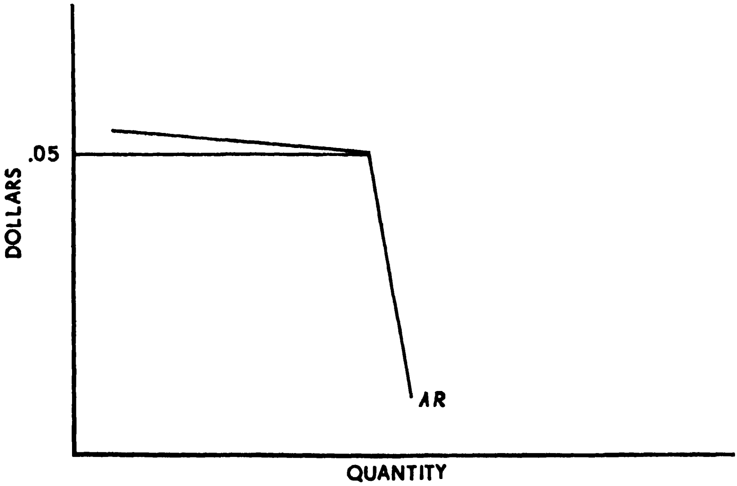
shaped demand curves.[3] Thus, the demand curve had substantially different elasticities at different points.

Customary Prices

Examples of customary prices are the 5-cent candy bar, chewing gum, soft drinks, and similar types of convenience goods the prices of which are largely a matter of tradition. For this reason the prices of such items have tended to persist, because management assumes a type of kinked demand curve as in Figure 8–4. At prices above the customary

FIGURE 8–4

Customary Price



---

[3] O. Knauth, "Some Reflections on Retail Prices," in *Economic Essays in Honor of Wesley Clair Mitchell*, pp. 203–4.

price, in this case 5 cents, sales fall off rapidly, evidencing high elasticity; at prices less than the customary price of 5 cents, sales increase less than proportionately, indicating relative inelasticity. The demand curve thus contains a kink at the customary price. Examples of this situation are not difficult to find. Most candy manufacturers, for instance, have chosen the alternative of reducing the size of candy bars in the face of inflation and higher costs since World War II, rather than alter the customary 5-cent price. In the public transportation industries, some firms preferred to postpone the costs of replacement and maintenance of equipment (e.g., streetcars) thereby permitting a deterioration in quality of service, rather than petition the transit commission for a rate increase.

Pricing at the Market

The kinked demand curve of Figure 8–4 also represents a type of policy sometimes referred to as "pricing at the market." It arises in instances where (1) management is ignorant of the true shape of the demand curve confronting it and hence adopts a "safe" policy of matching its price with competitors, or (2) where oligopoly markets prevail and a policy of matching competitors' prices minimizes the chance of a price war. In either case, consumers must regard product differentiation between competitors as basically insignificant if the kink is to exist. In that case, a small increase in price above the kink by any firm will bring it a large loss in sales, while a decrease in price will be followed by competitors and the resulting increase in sales will be relatively small for each. The frequent price wars between gasoline stations in many states are a typical example.

Prestige Pricing

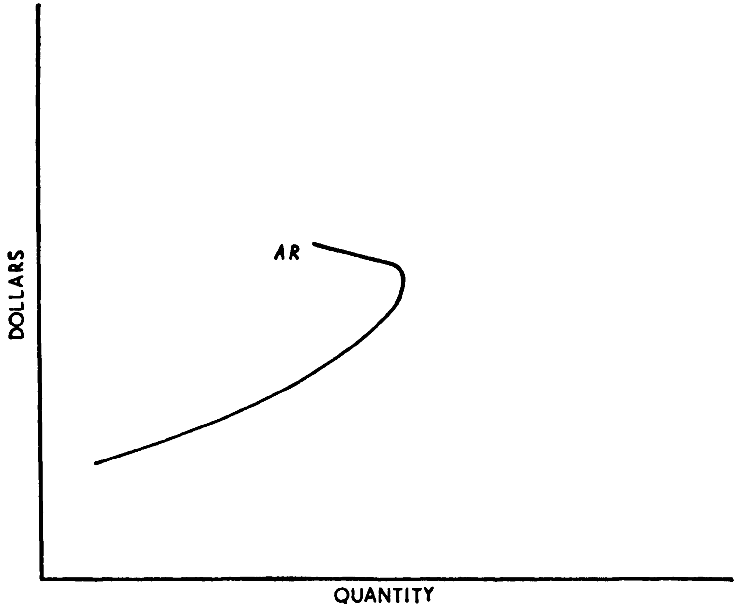
Where buyers judge the quality of a product by its price, the resulting demand curve will be positively inclined over a range of quantity, and may eventually bend back again as in Figure 8–5. The curve shows that a larger quantity is actually demanded at higher prices, until finally the price is sufficiently high that smaller quantities are demanded. From a theoretical standpoint, a positively inclined demand curve represents an extreme case of irrational consumer behavior.[4] Concrete examples may exist, however, for such luxury goods as fine furs, diamonds, and expensive trips, and even in the more everyday buying habits of consumers, at least in the Veblen sense of "conspicuous consumption."

---

[4] Some economists have contended that positively inclined demand curves are not a contradiction or exception to economic theory if they exist because consumers regard price as one of the product's qualities. In that case, they hold, the demand curve does not measure the same product, but rather "different" products along the same curve, and hence the demand curve is not the same demand curve of economic theory. In reply, it can be said that, from the seller's standpoint of pricing policy, what counts is what *he* believes is the shape of the *AR* curve for his product, irrespective of consumer psychology and its representation by indifference curves.

FIGURE 8-5

PRESTIGE PRICING



Price Lining

The policy of price lining, frequently found in retailing and usually practiced by department stores, refers to the offering of a class of merchandise in a limited number of price lines according to differences in workmanship, materials, design, or other characteristics. The classes of products to which the practice is often applied include coats, suits, dresses, furniture, hosiery, novelty jewelry, and a wide range of other items. Price lining is thus a manner of exploiting quality differentials. Once the lines are decided upon, they are usually held constant for long periods of time, with changes in market conditions adjusted for by changes in quality rather than price lines. Frequently, only three basic price lines are deemed necessary for each type of merchandise, on the assumption that the customer needs some basis of comparison before he can make a decision to buy at a particular price. The three price lines represent a "good," "better," and "best" plan, the lowest prices being for a stripped item and the medium and higher prices representing improved quality, styling, and other selling appeals. Sometimes each price line is actually a range of prices, called "price zones," according to differences in customer preferences.[5]

---

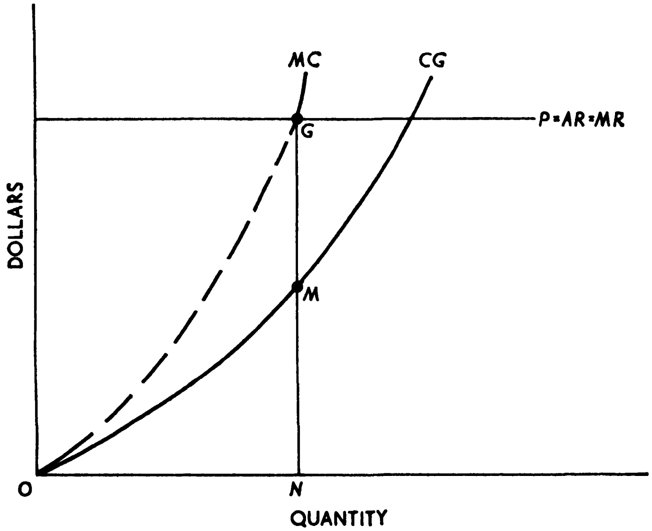
[5] See E. A. Filene, *The Model Stock Plan,* pp. 14–35, for a detailed defense of price lining; also, on this and other practical pricing issues in retailing, there is the article by Q. F. Walker who is economist for R. H. Macy & Co., Inc., "Some Principles of Department Store Pricing," *Journal of Marketing* (January, 1950).

Two chief advantages claimed for price lining are that: (1) it simplifies the price structure, thereby enabling manufacturing and selling effort to be concentrated on the most profitable price lines; (2) it avoids the need for making frequent pricing decisions after the establishment of the initial price and with the exception of special sales.

With respect to the first argument, in periods of rising manufacturing and selling costs, quality may have to be reduced to preserve customary price lines, or else frills may be added to the product to raise it into the next higher price line. If the retailer has heavily promoted a particular price, he may have to accept a lower margin and/or reduce his quality in order to maintain his advertised price. In periods of declining business activity, wholesale and retail prices tend to be "sticky" because better materials and workmanship are added to preserve the higher price lines of prosperity times. Price flexibility is thus lessened, creating an obstacle to the dynamic sort of pricing that would be more advantageous both to the company and to the economy as a whole over the long run.

FIGURE 8–6

PRICE LINING



Concerning the second argument, a price lining policy does not avoid management's problem of making price decisions. In fact, it presents the seller with precisely the same choice of alternatives as does a variable price policy, namely, whether to price by (*a*) equating marginal cost with marginal revenue, or (*b*) using a customary per cent of markup. The decision, however, concerns the prices paid for merchandise rather than the selling prices. The widespread use of price lining by retailers has resulted in manufacturers and wholesalers giving increasing attention to tailoring their own prices in order to fit retail prices. The retailer, however, does have some choice with respect to the quality of goods he purchases, and, presumably, the more he pays or the lower his per cent of markup, the larger his sales volume must be to yield a given profit. Thus, in Figure 8–6, if *P* is the established retail price, it is also the average revenue and marginal revenue curve since the line is horizontal. The line *CG* represents the cost of goods in various quantities that can be bought by the retailer. Evidently, the retailer should equate *MC* with *MR* (the price), paying *NM* for the merchandise and selling the quantity *ON*, thereby obtaining the maximum gross margin, *GM*. Alternatively, he may buy at a price that provides a customary or arbitrary per cent of markup, but then it would be a matter of pure chance as to whether he obtained the maximum gross margin.

Actually, other than cost of goods, there are relatively few variable costs associated with sales at retail. The retailer's goal, therefore, should be generally one of maximizing *gross margin dollars*. To the extent that other variable expenses are significant, however, they may be added to cost of goods, and the *CG* curve in Figure 8–6 would then become an average variable cost curve instead.

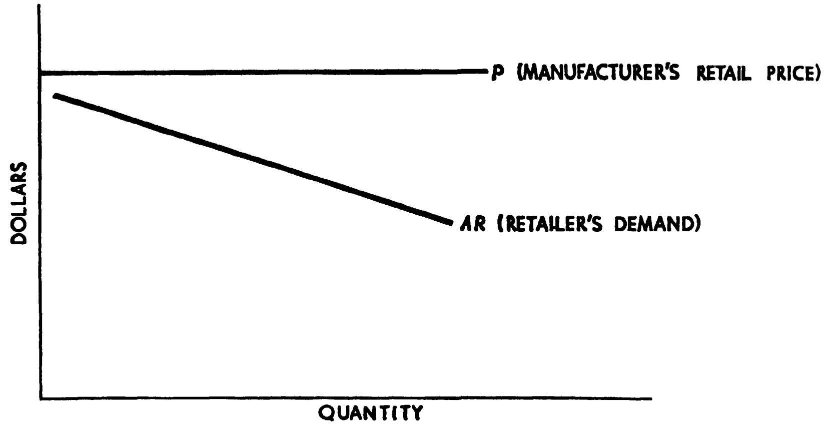
Resale Price Maintenance

Resale price maintenance is a form of vertical price control. It occurs when a price agreement is made between two sellers at different levels in the distribution channel, such as a manufacturer and wholesaler, manufacturer and retailer, or wholesaler and retailer, whereby the minimum or actual wholesale or retail prices of a product bearing the producer's trademark, brand, or name are fixed by contract. The states that have on various occasions upheld such contracts have done so under what is commonly referred to as "fair-trade" laws or "unfair practices" acts.

From our present standpoint, resale price maintenance is interesting because the assumption is frequently made that the retailer, for example, has no pricing decisions to make when a manufacturer maintains resale prices under fair-trade contracts. Actually, the retailer in this instance will always have at least one decision to make and possibly a second as well. He must choose between a pricing policy that equates marginal costs and marginal revenue as against one that employs a customary per cent of markup. A selection of the latter may well result in his refusal to push or even to handle many low-markup items that actually would be very profitable to him. In those states where the fair-trade laws permit only minimum rather than specified prices, the retailer must choose between selling at the minimum or some higher price. Selecting the latter would involve still further decisions as to what price policy to adopt.

What is the nature of the pricing decision for a *manufacturer* using resale price maintenance? Figure 8–7 provides an illustration. At any given retail price, *P*, set by the manufacturer, he will be confronted with a particular demand curve by retailers, *AR*, the shape of which will depend on their attitudes concerning the amount of markup they can obtain on the manufacturer's selling price. If the markup is low, some retailers will refuse to take the item and others will refrain from pushing it. If the markup is relatively high, dealers will tend to push the item and hence sell more than consumers would otherwise have taken at the given retail price. The appropriate price policy of the manufacturer, therefore, is to establish his optimum price by first computing his marginal revenue from this average revenue, and then equating his marginal revenue with his marginal cost. Since there will be a different *AR* curve associated with each retail price, the calculations must be made for each retail price, and the combination of retail and wholesale prices that will yield him maximum profits is then selected.

FIGURE 8–7
RESALE PRICE MAINTENANCE



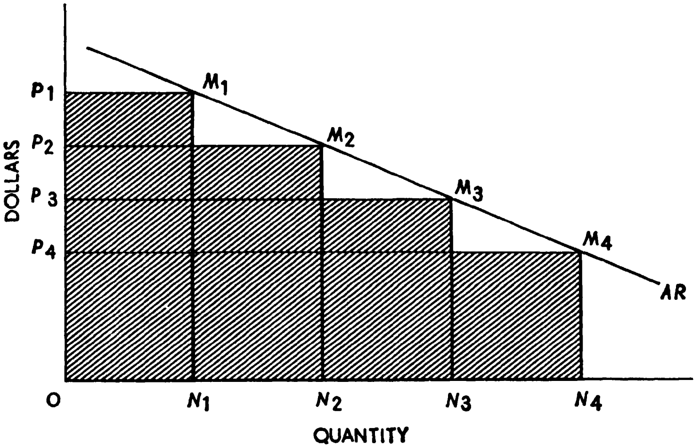
Quantity Discounts

In marketing literature and in statements made by businessmen, quantity discounts are usually justified in terms of: (1) the lower *unit* costs of handling larger orders because certain costs remain fairly constant or else increase less than proportionately to the increased volume (e.g., bookkeeping costs), and (2) the desire to utilize excess capacity thereby further reducing unit costs. What type of price policy should a seller adopt if he is to offer quantity discounts to buyers? As before, the rules can be derived from the construction of simple model situations. Two types of quantity discount problems can be examined: the first, where the seller offers the product to the *same buyer* at quantity discounts; the second, where the seller offers the product to *different buyers* at quantity discounts. As always, the aim of the analysis is to find out if theory has anything to offer as a guide in establishing an optimum price policy.

Figure 8–8 illustrates the process of quantity discounting to one buyer. Assuming *AR* to be the buyer's demand curve, the seller may simply charge a price of OP_4 per unit and thereby sell ON_4 units. Since there is no discount involved, his total revenue is then the area of the rectangle $OP_4M_4N_4$. However, he can enlarge his total receipts considerably if he offers quantity discounts. Thus, he may first charge a price of OP_1 and thereby sell only ON_1 units, giving a total revenue of $OP_1M_1N_1$. Then he may lower his price to OP_2 per unit and sell an additional N_1N_2 units, then lower the price to OP_3 and sell an additional N_2N_3 units, and finally lower it to OP_4 where he sells a further N_3N_4 units. Although he still ends up selling the same number of units, namely ON_4, his total receipts are now the entire shaded area instead of the area $OP_4M_4N_4$ when he charged a price of OP_4 per unit without quantity discounts. Evidently, the smaller he can shade his discounts, the narrower the steps under the demand curve become and hence the larger his total revenue. Theoretically, the limit would be a total revenue equal to the entire area under the curve, but this would require discounting in infinitesimal amounts. In practice, of course, the discounts are in blocks of units.

FIGURE 8-8

QUANTITY DISCOUNTS: ONE BUYER



Examples of this type of pricing, called "differential pricing" and discussed again later in this chapter, are not uncommon. They illustrate not only a type of quantity discounting, but may also account for why a surgeon may charge a rich patient more than a poor one for the same operation, and why a theater may charge more for some seats than for others for the same performance. In all instances, the seller is assuming a downward-sloping demand curve and is attempting to tap it in segments by chiseling away at successive portions of the curve.[6]

---

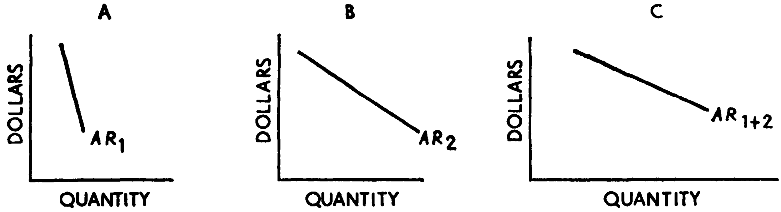
[6] Electric utility companies are an outstanding example, charging decreasing rates in blocks of killowatt hours, or perhaps establishing different price scales for industrial and residential users.

Figure 8–9 illustrates the process of quantity discounting to more than one buyer. Although only two buyers are represented here for purposes of simplicity, the analysis can be generalized so as to include any number of purchasers. Figure 8–9A shows a buyer whose demand curve, AR_1, is relatively inelastic, whereas Figure 8–9B shows a buyer whose demand curve, AR_2, is relatively elastic. Evidently, it would be foolish to offer both buyers the same discount schedule, since the first buyer, because of his inelastic demand, would take almost as much of the product at high prices as at low ones. The second buyer, however, at lower prices, would take more than proportionately larger quantities, as shown by his elastic demand curve. Assuming that the buyers can be sealed off in separate markets (as discussed later in this chapter), the correct procedure would be to offer each a quantity discount schedule according to his respective demand curve, as illustrated previously in Figure 8–8. The total demand confronting the seller would then be the curve in Figure 8–9B, derived by summing the quantities that would be taken by each buyer at a given price. That is, AR_{1+2} is the summation of the horizontal ordinates of Figures 8–9A and B. In general, the correct pricing policy will be the one that charges a higher price scale to the buyer with a less elastic demand (Fig. 8–9A), and a lower price scale to the one with the more elastic demand (Fig. 8–9B). If the demand elasticities are the same for both buyers, it will pay best to charge both the same price scale.[7] The process thus consists of tailoring the quantity discount schedule to each individual demand curve. Sellers who construct quantity discount schedules with an eye to their effects on certain large buyers are probably attempting to accomplish precisely these ends, although in an approximate manner by trial-and-error methods rather than by measurement based on guiding principles of economic theory.

FIGURE 8–9

QUANTITY DISCOUNTS: MULTIPLE BUYERS

A, Inelastic Demand; *B*, Elastic Demand; *C*, Total Demand



Geographic Pricing

Finally, a comment on the spatial aspects of pricing is in order before proceeding to the problem areas discussed in the remaining sections of this chapter. The problem of establishing an optimum geographic pricing policy revolves largely around the existence of transportation costs and certain legal considerations. Postponing the more detailed aspects of these factors for later discussion in this and the following chapter, some essential points may be noted at this time.

---

[7] Where discounts are not involved, the condition for maximum profit is that the seller equate the marginal cost of his entire output with the marginal revenue from each buyer. This is a common form of price discrimination, as discussed in the literature of economic theory.

The appropriate price policy is *not* simply one that charges all buyers the same base price with the result that buyers closer to the plant pay less and buyers farther away pay more, according to differences in transportation costs. Instead, each buyer's average revenue curve should be conceived as a *net* demand curve after deduction of transportation costs. The elasticity of each buyer's curve then becomes the important factor, in that the correct net price to each buyer is the one that equates the marginal cost of the seller's entire output with marginal revenue.

Under the Robinson-Patman Act, a seller is not allowed complete freedom of differential price discretion. Thus, price differentials must not be larger than cost differentials. However, some discretion does exist. For example, the seller may, at least within limits, employ a differential pricing policy: (1) by giving discounts that are less than the amount of his cost saving; (2) when buyers are not in competition with one another; (3) where he is himself "meeting competition" but not "beating" it; and (4) by selling slightly different products under different brand names. Thus, it appears that economic theory does offer a guide for management decisions involving geographic pricing policies, as will be seen more fully later on.

Conclusion

This section has had two main purposes: (1) to provide a theoretical background against which to formulate intelligent pricing policies, and (2) to show that most pricing behavior, which is treated in a purely descriptive way in virtually all marketing discussions, can actually be integrated with the general theory of monopolistic competition. This second consideration has two particular advantages. First, it enables managerial economists to learn more about the pricing policies actually used by business firms and hence provides a sounder means for proposing improvements in these policies. Second, it provides a guide for empirical measurement by focusing sharply on what it is that really needs to be measured. The result is thus a stronger foundation on which to construct a plan for future action by management. This will subsequently become more evident as many of the pricing concepts discussed only briefly here are treated more fully in the sections that follow.

PRICING METHODS

Against the background of pricing theory outlined in the previous section, we turn our attention to a survey of alternative pricing methods frequently employed in industry. Only the most common methods will be discussed, with the exception of differential pricing, which is important enough to require a separate section later in this chapter. The pricing methods considered are: (1) cost-plus pricing, (2) flexible markup pricing, (3) intuitive pricing, (4) experimental pricing, and (5) stable and imitative pricing.

Cost-Plus Pricing

The most widely used method of pricing employed by business firms is known as cost-plus pricing. It is a procedure whereby the price is determined by adding a fixed markup of some kind to the cost of the good (as distinguished from a variable markup, which is discussed later under "flexible markup pricing"). Thus, a manufacturer pricing by a cost-plus method, if he desires a 20 per cent markup, would price at $10.00 a good whose cost was $8.00. Evidently, two areas of uncertainty, and hence two decision problems, confront the manager who uses cost-plus formula pricing: (1) arriving at an estimate of cost, and (2) selecting the appropriate margin or markup. How are these done by most firms?

In practice, most manufacturers using cost-plus pricing usually employ some notion of standard cost as their basic cost figure. They arrive at the figure by estimating unit costs of labor and materials and by computing unit overhead costs for operations at some arbitrary percentage of capacity. In other words, they typically calculate their costs for a "standard output," commonly between two thirds and three fourths of capacity, irrespective of the actual volume of operations. Other cost measures sometimes used, however, are actual cost, or the cost for the most recent accounting period, and expected cost, which is a forecast of actual cost for the future pricing period based on a forecast of operating rates for that period. Still another method, but one that is relatively rare in industry except where special products are concerned, is to construct a cost figure based on engineering estimates of efficiency and various physical relationships. In any case, regardless of the method employed to estimate costs, the over-all nature of the pricing formula is essentially the same.

Numerous surveys of pricing practices have been made by economists, and hundreds of pricing methods were filed with the OPA during World War II. On the basis of all this information, the evidence seems to indicate no definite answer to the question of how the size of the markup is decided upon, other than the feeling on the part of businessmen that their margins represent what they believe to be a "fair" or "reasonable" profit. The evidence also indicates that there are wide variations in the percentage of markup both within industries and between industries, due to differences in competition, cost structures, accounting methods, inventory turnover, and custom. This last factor, custom, appears to be of considerable significance. Margins used in the past are considered "fair" simply because of their long use over the years. Scarcely a businessman surveyed believes that the margin used is the most profitable one; all seem to stress the ethics rather than the economics of price setting; and most if not all are aware that a more profitable pricing policy may be possible.[8]

What are the reasons given by businessmen for the wide prevalence of cost-plus pricing? Some of the chief ones are these. (1) It offers a relatively simple and expedient method of setting price by the mechanical application of a formula. (2) It provides a method for obtaining adequate ("fair") profits when demand is unknown. (3) It is a method of establishing a stable price uninfluenced by fluctuations in demand, which is particularly important to firms that commit themselves on price through their catalogs, advertising, etc. (4) It is desirable for public relations purposes even at the expense of short-run profits, presumably because consumers will accept price increases when costs rise and will expect price reductions when costs decline. Failure to pass along a known cost decrease in the form of a price reduction may cause consumer antagonism and a shift in patronage. Recognition of this prompted General Motors in the postwar period to reduce its car prices in accordance with a downward wage and salary adjustment, even though dealers had backlogs of orders and GM cars were already selling at lower prices than comparable cars of other makers.

Despite its prevalence, cost-plus formula pricing has at least three important disadvantages to firms employing it as a pricing method.

1. It fails to take account of demand as measured in terms of buyers' desire and purchasing power. Moreover, where price planning for the future is involved, what is needed is a forecast of both future costs and future demand if the best pricing decision is to be made, and not an estimate of past or even of present costs.

2. It attempts to make an accurate measure of what usually amounts to the wrong cost concept, rather than even an approximate measure of the right cost concept. What is frequently needed, for example, is at least a rough estimate of opportunity costs (sacrificed alternatives) and of incremental costs, neither of which are readily available from accounting records, rather than accurate estimates of irrelevant concepts such as past or present costs. Further, there is even some doubt as to how accurate the measures of full (past or present) costs usually used in the formula really are, especially in multiple-product firms where common costs exist and are hardly more than arbitrarily apportioned to products in typical cost accounting systems.

3. It fails to reflect competition in terms of rivals' reactions and the possible entry of new firms. For example, in an industry that prices by the cost-plus method, if company margins are above the level necessary to cover operating costs and yield "normal profits" per unit at capacity, new firms will tend to enter the industry as long as no considerable excess capacity is already present. The result will be a smaller market share

---

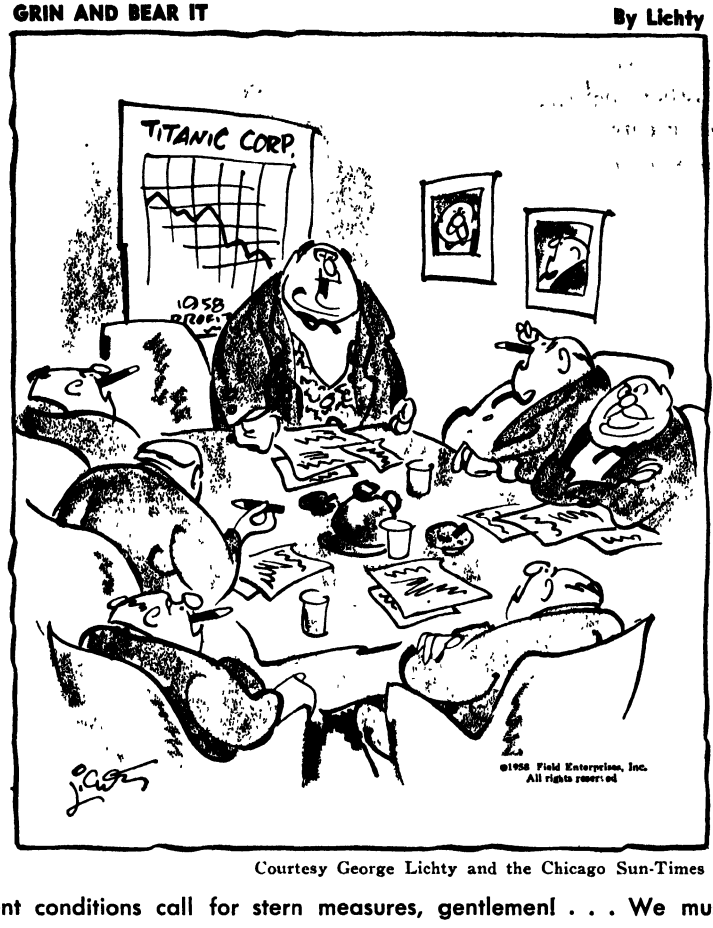
[8] Some better-known surveys of pricing practices are listed in the bibliographical note at the end of the chapter.

for each firm, and therefore higher unit overhead costs and lower profits per firm.

In view of the above considerations, it appears that cost-plus formula pricing is justifiable by management as long as demand elasticity and the industry's competitive structure are not known, and as long as stockholders are content with adequate profits rather than maximum profits.

Flexible or Variable Markup Pricing

A pricing practice that is closely related to cost-plus pricing, but is by no means as widespread in industry, has been termed flexible or variable markup pricing. Essentially, it is a pricing method that takes some cognizance of changing economic conditions by providing for a variable markup over the course of a business cycle. In periods of prosperity when incomes are high and buyers are less price conscious, sellers add larger



Courtesy George Lichty and the Chicago Sun-Times Syndicate

"Present conditions call for stern measures, gentlemen! . . . We must slash our next price increase by 10%! . . ."

margins to their base cost; in recession or in relatively low income periods, buyers are more sensitive to competitive price differences, so sellers add smaller margins to their base cost.

Flexible markup pricing, despite its advantage over cost-plus pricing, in that at least it takes some recognition of demand, is not a common pricing practice used by business firms.

1. It requires frequent estimates of demand which ordinarily involve more time, effort, and money than most industrialists care to expend. Besides, there is a common and sometimes well-founded belief among many managements that the longer-run sale of their products is more affected by changes in determinants other than price, such as incomes, advertising, and the prices of substitutes where consumer durables are concerned, and buyers' profit anticipations where durable producers' goods are involved. In addition, buyers of consumer durable goods often react to major price cuts not by immediate purchase increases, but by postponing their purchases in anticipation of still further price reductions. All of these factors tend to indicate that many sellers frequently regard the cyclical demand for their products as relatively price inelastic, which probably accounts to a large extent for the more widespread use of cost-plus rather than flexible markup pricing.

2. Sellers will tend to prefer cost-plus pricing to flexible markup pricing during recession periods, in the belief that they deserve a larger margin when business declines. When sales decrease, cost per unit (exclusive of merchandise) increases, and the seller sees his margin, which allows for overhead costs, eroding. For even if the margin percentage remains the same, the absolute amount declines because of the drop in base costs, and hence sellers feel that their charges to buyers are lower.

3. The goal of a "fair" price is lost if margins are allowed to vary, and instead takes on a flavor of "charging what the traffic will bear." The objective of "reasonable" profits is not uncommon and seems to be an important motive guiding many management decisions.

Intuitive Pricing

Intuitive pricing, or pricing by the "feel of the market," is a fairly common method practiced by many executives. The degree of its application can vary from prices based on pure hunches or guesses to prices based on an examination of past data and future trends in costs and demand. As a price-making method, it bears in many ways the same relation to pricing practices and policies as does the factor-listing method to business forecasting. A common procedure in many firms is to arrive at a preliminary price estimate based on a cost-plus formula and then adjust the price upward or downward in accordance with executive opinion as to expected demand, competition, and other market forces. In a certain sense, therefore, it combines cost-plus pricing with flexible markup pricing. The emphasis, however, is on the subjective "weighting" of factors believed to be influential in affecting price, and thus is a type of psychological rather than mechanical pricing method. The following statement by Ernest Breech, executive vice-president of Ford, is fairly typical: " . . . such a price [based on a cost-plus formula] is obviously only a 'standard' in the sense of being used for comparative purposes. It is a useful guide to judgement . . . The final prices are, of course, determined by the competitive situation."[9] Evidently, intuitive pricing requires a high degree of self-confidence, since the firm's well-being will depend to a large extent on how accurately management can "feel" future business conditions. Despite its vagueness, a possible justification of the method is that the extreme subjectivism involved in this type of pricing often requires that the pricing decision be the result of group action, which thereby removes the responsibility of a wrong decision from the shoulders of any one executive.

Experimental Pricing

A technique for arriving at an optimum price that has gained increasing acceptance by companies in recent years is a kind of trial-and-error, or experimental, pricing. The procedure is to select a sample of test markets, establish an experimental (e.g., Latin square) design, and by manipulating the treatments as described in Chapter 3 on economic measurement, arrive at a price that maximizes profit. However, because of the difficulty of deriving empirically the price that will actually maximize profit, the more common practice is usually to choose the price that maximizes sales. Experimental pricing thus offers at least a partial solution to the problem of establishing an optimum price by taking some recognition of the influence of demand.

Experimental pricing methods have found particular application in the pricing of new products at the retail level. Conducted properly, these experiments can yield rich marketing information for later use as well as a sounder base on which to construct a more profitable pricing structure. As in researching demand, however, the approach through controlled experimentation can be hazardous, whether for new products or for established ones, when: (1) oligopoly conditions prevail in the test markets so that there is a danger of rivals' reactions to downward price movements by the experimenting firm, and (2) buyers cannot be sealed off into separate markets so as to prevent their infiltration from higher- to lower-priced markets. The first condition has been treated earlier and is examined again below with reference to stable pricing; the second is discussed later in this chapter in the section dealing with differential pricing.

Stable and Imitative Pricing: Oligopoly Problems

A company that adheres strictly to any one or combination of the above four pricing methods would, in view of the dynamic economic environment, be in almost a continuous process of rebuilding its price structure. The fact is, however, that many firms do not recalculate their price structures frequently, but instead establish prices either by building on stable prices of the recent past or by imitating the prices charged by competitors. The reasons behind each of these pricing methods may be examined briefly.

---

[9] From an address before the American Marketing Association, June 11, 1947, New York City.

Stable Pricing. Price stability for a period of months and sometimes for years is the rule for most companies rather than the exception. Official quotations in catalogs and other media, wage contracts, and product differentiation are only a few of the important factors making for price stability of manufactured goods. But in oligopolistic industries characterized by few sellers and very similar products, which is the typical situation in American manufacturing, there are further price-stabilizing influences as well.

1. As explained with reference to the kinked demand curve, sellers in industries with few firms will normally tend to maintain the prevailing price, since, in theory, competitors will not usually follow price increases but will match price reductions, at least in particular market areas. The effect of a price reduction by any one seller, therefore, is an increase in sales by all firms matching the price reduction, with no significant diversion of customers from one competitor to another and hence no gain in market share. Thus, a building materials producer states, in reply to a recent questionnaire: "If we don't keep up with our competitors' price reductions, our older customers usually call us and tell us what we must do to secure their business, and that usually amounts to matching the price. Sometimes our newer customers don't even call us first; they just buy elsewhere and our salesmen have to try to recoup the lost business the next time around."

2. Even aside from the oligopolistic aspects of the problem, there are less complex reasons why price stability is an important element of a firm's price policy. Essentially, changes in price can be costly to the company as well as disturbing to salesmen and purchasers. Many firms that reduced their prices in recession periods found it extremely difficult to raise them later, even in the face of rising costs and general inflation.

Imitative Pricing. Imitative pricing occurs when a firm chooses to set its price equal to, or at some proportion of, the price of another firm in the industry. The advantages to the imitator of pricing products this way are that: (1) the firm it imitates may be more experienced and better able to establish the appropriate price, (2) it saves the expense of deriving demand and cost estimates, and (3) it leaves management more time to concentrate on nonprice competitive forms, such as advertising, merchandising, product development, personal selling, and services. The significance of these conditions that make for imitative pricing is quite important and should not be underestimated. The tendency to rely on an experienced competitor in setting prices, the tendency to avoid price cutting because of its retaliatory effects, and the likelihood that nonprice competitive practices would, in the long run, increase industry demand more than would lower prices, have created a unique type of competition in oligopolistic industries. One aspect of this competitive pattern is a type of imitative pricing known as "price leadership," to which some special attention may be devoted at this time.

Price Leadership. When firms tend to establish their prices in a manner dependent upon the price charged by one of the firms in the industry, price leadership exists. The firm that takes the initiative in announcing its changes in price is called the price leader; all other firms in the industry that either match the leader's price or some differential of it are termed price followers. The price leader will usually be a leader in all markets, although it frequently happens too that a firm will sometimes follow in some markets and sometimes lead in others. Custom, industry demand and cost structures, and changing economic pressures vary between industries, and changes in these combinations are frequently sufficient to destabilize existing patterns for unpredictable periods of time.

Price leadership can easily exist without explicit agreements and, indeed, this form may well be the rule rather than the exception. That is, price leadership frequently arises as a natural growth within an industry, and the price leader is usually the firm with a successful profit history, sound management, significant market share, and long experience in marketing matters. The remaining firms in the industry accept the leader, not necessarily because of an explicit agreement, but because of his ability to coordinate the industry's growth with that of its members. In effect, the leader's over-all judgment of market conditions replaces the separate judgments of the followers. The procedure is well illustrated by the following passage from a TNEC hearing dealing with firms engaged in the fabrication of nonferrous alloys. The American Brass Company had a 25 per cent market share at the time (late 'thirties); Mr. H. L. Randall, President of the Riverside Metal Co., a small New Jersey firm with less than 2 per cent of the market, is testifying before the Temporary National Economic Committee:

Mr. Cox: Mr. Randall, would it be correct to say that there is a well crystallized practice of price leadership in the industry in which you are engaged?
Mr. Randall: I would say so.
Mr. Cox: And what company is the price leader?
Mr. Randall: I would say the American Brass Company holds that position.
Mr. Cox: And your company follows the prices which are announced by the American Brass?
Mr. Randall: That is correct.
Mr. Cox: So that when they reduce the price you have to reduce it too? Is that correct?
Mr. Randall: Well we don't have to, but we do.

Mr. Cox: And when they raise the price you raise the price?
Mr. Randall: That is correct.

. . . . .

Mr. Arnold: You exercise no individual judgment as to the price you charge for your product, then?
Mr. Randall: Well, I think that is about what it amounts to, yes sir.[10]

Mr. Randall also testified: "We follow the prices set by the bigger companies and pray that we will make a profit." In explaining why, he stated, "because it is the custom of the industry. We have always done it." Explaining further, he noted with reference to the American Brass Company, "I think they know what they are doing; they probably know what their costs are a lot better than I do. . . . I must confess that our costs are very sketchy. We have a cost department I think of three people."[11]

Price leadership is found in a great many industries other than nonferrous alloys. The same TNEC investigations brought out its existence, for example, in steel with U.S. Steel as the leader, and in glass containers with Hazel-Atlas leading in wide-mouth container ware, Owens-Illinois in proprietary and prescription ware, and Ball Brothers in fruit jars and jellies. Evidence of a leader-follower relationship also has existed in such industries as agricultural implements, cement, cigarettes, copper, gasoline, lead, newsprint, sulphur, and tin cans, to mention only a few. Not in all cases, it may be noted, do small firms only follow the leader down, in accordance with economic theory and the dictates of competition; in reality, they frequently follow a price increase as well—a fact not often emphasized in theoretical discussion but one that is readily observable in practice. The reasons may be: (1) a fear or desire on the part of the price follower to avoid provoking a price war with the leader, (2) a belief by the follower that profits are larger in the long-run under the refuge of the leader's price umbrella for the industry as a whole, or (3) merely because the follower finds it easier or more convenient to follow the leader. In any case, as long as the leader's price is high enough to allow at least normal profits for the less-efficient followers, the industry may operate fairly smoothly with little or no price warfare. The challenge to the leader may thus involve not only an ability to forecast changing demand and cost conditions, but sometimes also to construct an industry-wide pattern of price differentials that would be acceptable to members and that would allow for differences in brand name, service, and quality, particularly in differentiated oligopolistic industries. Failure to comply with these conditions and to revise the differential structure with changes in underlying market conditions and business cycles may easily result in a loss of leadership, despite the firm's historical dominance in the industry.[12]

---

[10] *Hearings* before the TNEC, Part V, pp. 2085–87.

[11] *Ibid.*, pp. 2098–99.

[12] Recent developments in the automobile industry are an example. With Ford —and to a lesser extent, Chrysler—approaching General Motors in cost efficiency, the

Summary and Conclusion

Businessmen employ a variety of pricing methods of which cost-plus formula pricing, because of its simplicity and mechanical nature, is most common. Though these methods frequently provide adequate profits, they do not provide maximum profits because they ignore demand in general and its elasticity in particular. Perhaps the most important factor coloring managerial price decisions are businessman's notions of "fair" or "reasonable" profits as a goal of the firm. It is quite possible, however, indeed even likely, that the concept of a just price is a rationalization by executives to compensate for their economic ignorance rather than a means of securing only a moderate profit as the company's objective. If a management could, on the basis of well-founded demand and cost calculations, estimate the most profitable price consistent with other company objectives, it seems plausible that competitive forces would prompt it to charge that price at least in the long run. To the firm that knows its costs and demand, therefore, the notion of a fair profit can readily be reconciled with that of maximum profit; to the firm that is ignorant of its demand, costs, and market structure, however, any profit—even if not a maximum—would obviously be "fair." Evidently, there is a need for serious reconsideration on the part of many managements as to the profitability of their present and usually outdated pricing practices and policies.

PRODUCT-LINE PRICING

Most of economic theory with respect to pricing is based on the assumption that the firm produces only one product. If "product" is defined broadly—for example, to mean automobiles, men's shoes, or locomotives—this assumption is not unrealistic and, in fact, goes a long way in describing a very important part of business behavior. But for many management problems such a broad definition is unsatisfactory because it fails to explain why a firm produces diverse commodities and what the relationship is among their prices. In modern industry the typical firm produces multiple products, and therefore a definition of "product" is needed that is more suitable for attacking the kind of pricing problems encountered by such firms. The most meaningful definition is simply that *a product is any homogeneous commodity*. But what is the criterion of homogeneity? In everyday terms, the test is that buyers must not distinguish between any portions of the stock, or, in other words, that they

---

latter's historical leadership is seriously questioned. Thus Ford and Chrysler, by 1958, were getting into a position to set prices without regard for those of General Motors or any other competitor—a fact which was not always true in the industry. (See "Has GM Lost Price Leadership?" *Business Week* (November 9, 1957), p. 171.)

be indifferent as to separate portions.[13] Thus the key economic feature with respect to the pricing of a company's product line is the nature of the interrelated demands for parts of the firm's output, which when measured in quantitative terms takes its most common form as the cross-elasticity of demand. As noted in an earlier chapter, this coefficient represents the percentage change in the demand for product Y resulting from a 1 per cent change in the price of product X, the price of Y remaining constant. On this narrow definition of product, it is evident that virtually all firms are multiple-product firms and that the emphasis from a pricing standpoint is to define a company's product line in terms of demand interrelationships.

Some production and cost considerations of firms producing multiple products were outlined in an earlier chapter on production management under the heading of "Product-Line Policy." It was noted there that a firm produces multiple products either because: (1) the demands for the various products are related, or (2) production costs are lower when products are jointly produced. Keeping this dichotomy in mind, some implications for product-line pricing may be noted under the separate conditions where goods are substitutes for each other and where they are complements. The over-all problem in product-line pricing is to manipulate the combination of prices until the optimum or most profitable price structure is achieved.

Pricing Substitute Goods

From the standpoint of product-line pricing, the production of substitute goods by a firm should be viewed as an effort to segregate (or segment) individuals or market sectors with different demand elasticities, in order to profit from the different taste idiosyncrasies. Striking examples of firms producing competing, i.e., substitute products, are numerous: meat packers, automobile manufacturers, tire companies, clothing producers, cigarette firms, soap companies, and pharmaceutical houses are only a few. Evidently, these firms compete with themselves to some extent in the sense that they produce products to fill similar needs. The more they sell of one product, chances are the less they sell of others. How then should these products be priced? In practice, two common methods of product-line pricing for substitute goods can be distinguished.

The procedure followed by most producers is to set prices on their entire line of products by the same method. Essentially, a markup method of pricing is used on the entire line of products, with the same margin employed for all similar products in the line. The specific technique is to price the products in proportion to costs, with the choice of costs being either full costs or transformation costs, the latter representing the labor and overhead expenditures required to transform (convert) raw materials into finished products.

---

[13] More technically and in terms of economic theory, the indifference curves between the two portions are straight lines. (See M. H. Spencer, "Demand Analysis: Indifference Curves," in A. D. Gayer, Harriss, and Spencer, *Basic Economics*, pp. 91 ff. for a brief explanation.)

A second approach commonly employed in product-line pricing is to price the product by varying the size of the margin with the absolute size of costs. Thus, the more costly the product, the higher the margin, and hence the higher the price.

Both of these methods, despite their widespread use in industry, suffer from the shortcomings that they take no account of differences in demand, differences in competitive conditions, and differences in the degree of market maturity of each product in the line. Further, the accounting methods employed to divide joint costs among products of the same firm are not at all justified economically, being wholly arbitrary and thus resulting in prices that reflect at least partly the arbitrary allocation of common costs. What, then, should be the appropriate method for setting price? Ideally, the optimum price in a market sector is the one that yields the largest contribution margin, tempered by expected secular shifts in demand, and by competitive forces as measured, for instance, by market share, the possibility of entry by new competitors, and other criteria of competitive intensity that may be selected as guides to action. Approached in this way, the product-line price structure would aim at the correct objective: that of exploiting the differences in demand elasticities between market sectors, with the maximization of future profits rather than current profits as the ultimate result. Management policy would thus recognize the future rather than the present as a basis for pricing decisions.

Pricing Complementary Goods

The second type of demand interrelation is complementarity. The degree of complementarity in use can take one near-extreme form of fixed proportions (e.g., wrist watches and watch bands, automobiles and engine blocks, houses and oil burners); it may take different degrees of variable proportions (e.g., turpentine and paint, cameras and film, hi-fi phonographs and records); or it may take the most remote form where the various products in the line are not jointly related in use but merely augment the firm's general reputation (e.g., dentifrices and soap by a firm such as Proctor and Gamble, where the ultimate product being sold is personal hygiene). In the last case, all multiple products of a firm can be viewed as complementary if they enhance one another's acceptability, but in any event the fundamental pricing principles are not materially altered. The ultimate objective, as with substitute goods, is to arrive at a price structure that produces the largest contribution margin according to the separate demand elasticities of market segments. An essential difference, however, is this: where complementary goods are concerned, a decrease in the price of one leads to an increased demand for the other, so that the cross-elasticity is significantly (in a statistical sense) negative. The direct price elasticity of demand would then be less than unity or inelastic. The practical consequence of this is that sellers will frequently find it more profitable to price an item low or even at a loss, in the hope of selling the complementary item at an above-average margin. Some illustrations of this and of similar pricing strategies for complementary goods, which are frequently encountered in descriptive form in marketing literature, may be noted as follows.

Loss Leaders. Loss leaders illustrate one type of product-line pricing of complementary goods. Most commonly encountered in retailing, this practice refers to the sale of one commodity at less than invoice cost or at a price sharply below customary price, and publicizing of the fact through advertising. The intention is: (1) to draw in customers who will buy other products, and/or (2) to arouse consumer interest that will eventually shift the demand curve to the right. In the first case the complementarity is between different products at the same time, and the direct losses on the loss leader are unimportant if they more than offset the indirect gains in the complementary items. In the second case, the complementarity reveals a time dimension between present and future demand, with the hope that present losses will encourage future sales and profits, e.g., magazine trial subscriptions, student rates on theater tickets, etc. For a loss leader to be effective, the cross-elasticity coefficient between the loss item and the other products must be large (ideally, infinite[14]); the direct or price elasticity should be low (ideally, zero); and the supply should be high so that the direct losses do not merely outweigh the indirect complementary gains.[15] Further, the good should be well known, widely and frequently purchased, unsuitable for storage by consumers, and standardized so that its customary price is widely known and "bargain" prices are quickly recognized. Thus, the phrase *loss leader* is actually a misnomer, for an intelligent management can in reality increase its profits by careful selection and pricing of loss leaders. Given the prices of other products, a change in the price of the loss leader produces larger sales of all products so that the increment in revenues exceeds the increment in costs. Therefore, a good loss leader is always a "profit leader."[16]

---

[14] And, of course, negative, except for prestige items as discussed earlier in this chapter, in the range where the demand curve is positively sloped. See Figure 8–5 above.

[15] Frequently, the purchasers of loss leaders will be rationed (e.g., one to a customer) at the submerged price. Implicitly, this serves to reduce the demand elasticity and limits the direct losses suffered by the seller while still evoking sales on the complementary commodities. (*Cf.* Roos, *Dynamic Economics*, for the mathematical construct, or Weintraub, *Price Theory*, chap. 14, for a further analysis of this and the two examples that follow.)

[16] It should be noted that, ignoring other effects, the fact that the cost of the leader is greater than its marginal revenue is irrelevant; the true marginal revenue of the leader is the change in the firm's total revenue with other outputs (or prices) re-

Tie-in Sales. Tie-in sales or contracts afford a second concrete illustration of complementarity commonly discussed in marketing literature. The practice consists of requiring buyers to combine other purchases with the featured goods, so that in effect the seller is offering the purchaser a joint product. Normally, the featured or "lever" commodity, if the tie-in is to be effective, must be difficult to substitute, not easily dispensed with, and relatively more inelastic in demand than the subsidiary item. An ideal opportunity for tie-in sales exists when the seller possesses an exclusive and essential patent, as in the classic example of the American Shoe Machinery Co. which compelled shoemakers to purchase other materials and intermediate products as a condition of purchasing shoe machinery. A variation of tie-in sales is "full-line forcing," where the dealer must accept the parent firm's entire product line as a condition of purchasing one item in the line. From the seller's standpoint, this may effectively seal off or at least curb sharply the distributive facilities of competing producers because of the limited financial and physical facilities of dealers. It also has a welfare effect, however, in that it reduces competition in distribution and narrows the alternatives open to consumers, thereby involving certain antitrust issues as to its legality, which are discussed further in the next chapter. In any event, the pricing aspect involves a recognition of the relatively inelastic demand for the main product and hence a higher price, coupled with a lower price for the subsidiary items because of their greater demand elasticity. Packaged sales, with the offer to "buy one and get one free" may be considered a type of tie-in practice, and is commonly encountered as a method of introducing a new product.

Two-Part Tariff. Still another illustration of complementarity in pricing and somewhat similar to tie-in sales is the "two-part" tariff. Here the buyer pays two prices for a joint product consisting of a fixed and a variable component. For the fixed portion, the buyer pays a set price independent of utilization, and for the valuable flow of services he makes separate payments. Examples include the basic installation charge for electric wiring or gas transmission lines and the variable payments dependent upon use; the minimum charge for public utility services and the variable payments for units purchased; the entry fee to an amusement park (or the cover charge in a night club) and the variable payments for each individual entertainment; and so on. Economically, the two-part tariff can be used as a device to cover initial costs with further income to be derived from the variable service (e.g., college registration fees and separate course rates per credit hour) or as a source of profit from both components of the product. In the latter case the product should be viewed as consisting of complementary items in variable proportions, with a rela-

---

maining constant (See G. Stigler, *The Theory of Price,* 1st ed., Chapter 16, for the general theory, and particularly pp. 312 ff. for diagrammatic techniques in multiple-product theory. Also, Weintraub, Chapter 14.)

tively inelastic demand for the fixed component and a more elastic demand for the variable flow.

Conclusions

The problems of product-line pricing are essentially twofold: (1) to decide on what it is that management wants and can expect from a structure of product prices, and (2) to manipulate the price structure until the desired combination of prices for producing the desired end is achieved. If the objective is maximum profits, it must be recognized that each price structure will produce a different sales mixture and therefore a different combination of total revenue and total cost. Hence the optimum price structure is the one that produces the greatest expected difference between total revenue and total cost, or in other words the largest expected net profit. To achieve this goal requires a knowledge of product interrelationships, particularly from the demand standpoint, since it is the demand elasticities rather than cost considerations alone that are usually more relevant for pricing purposes. The significance of cost, however, should not be underestimated. For instance, cost estimates are important when decisions must be made as to: (1) whether to drop the product or retain it, in which case the cost saving (i.e., avoidable cost) is relevant; (2) whether to charge this price or that, in which case a comparison between each product's differential (incremental) cost and price is needed, and when combined with sales forecasts, indicates the contribution profit for each product; and (3) whether to accept a sales commitment (e.g., government contract) for a fixed future period or quantity of supply, in which case the relevant criterion is again incremental cost, the size of which will depend to a large extent on future variations in capacity due to seasonal and cyclical factors. The goal, it should be emphasized again, is to manipulate prices so as to arrive at the *maximum expected total contribution to overhead and profits of all products combined.* Hence, both cost and demand estimates are needed.

Product-line pricing is thus seen to be closely related to problems of product-line policy as discussed earlier from the production standpoint in Chapter 6. One of the chief differences, however, is this: from the production viewpoint, the problem is one of manipulating the component items of the product line in order to maximize profit; from the pricing viewpoint, the problem is to manipulate the price structure of the component items in order to maximize profit. Both independent variables—the product line and its price structure—are necessarily interrelated in their effects on the dependent variable, net profit. The role of management in this respect, therefore, is to keep a continuous weather eye open for possible additions and deletions to the commodity belt if it is to maintain an optimum product line.

Finally, a word as to the role of product-line pricing is in order. Actually, the subject is closely related in many respects to that of differential pricing, as will be seen in the following section. Stress was placed, it will be recalled, on the segmentation of markets and on the significance of pricing substitute and complementary goods. Where products are neither substitutes nor complements, or only faintly substitutable or complementary, as denoted by cross-elasticities close to zero, they are treated as independent goods for pricing purposes, but the importance of demand and cost estimates are still applicable. In the next section, the significance of market segmentation becomes even more evident, and product-line pricing may be viewed as a basis for that discussion.

DIFFERENTIAL PRICING

Differential pricing (which may be regarded in certain respects as a form of product-line pricing) has been a subject of heated controversy for years, involving both economic implications and political repercussions. The economic aspects will be the main area of our concern in this section, with the political considerations postponed for the most part until the following chapter.

What is meant by the term "differential pricing"? Generally, it is a method that can be used by some sellers to tailor their prices to the specific purchasing situations or circumstances of the buyer. Specifically, it may be defined as the practice by a seller of charging different prices to the same or to different buyers for the same good, without corresponding differences in cost. For analytical purposes, it is convenient to distinguish between three classes of differential pricing.

First Degree. Differential pricing of the first degree means that the seller charges the same buyer a different price for each unit bought, thereby extracting the maximum total receipts. This was the type of pricing technique illustrated earlier in the chapter (Figure 8–8), with reference, however, to quantity discounts. By shading the price down to each buyer for each unit purchased, the seller obtains a larger total revenue than if he were to charge the same price per unit for all units bought.

Second Degree. Differential pricing of the second degree has the same underlying principle as first-degree pricing, except that the seller charges different prices for blocks of units instead of for individual units. The result is the "stair-step" pricing effect shown in Figure 8–8. The yield is still a larger total revenue to the seller than if he charged the same price, but not as large as would be realized if the price could be shaded for each unit so that the total receipts approached in magnitude the entire area under the demand curve (as in first-degree pricing).

Third Degree. Differential pricing of the third degree occurs when the seller segregates buyers according to income, geographic location, individual tastes, kinds of uses for the product, or other criteria, and charges different prices to each group or market despite equivalent costs in serving them. Thus, as long as the demand elasticities among different buyers are unequal, it will be profitable to the seller to group the buyers into separate classes according to elasticity, and charge each class a separate price. This is what has been referred to more generally in earlier discussions as market segmentation, i.e., the carving up of a total market into homogeneous subgroups according to some economic criterion. From the standpoint of pricing, the criterion usually employed is that of demand elasticity, and it is often applied in a practical manner via certain indirect means, as will be seen shortly.

The application of differential pricing to practical situations is based largely on the analytical tools developed in the economic theory of price discrimination. Hence, in this section, as well as in the following chapter, the two terms *differential pricing* and *price discrimination* may be regarded as synonomous unless the particular context in which the expression is employed indicates otherwise. With this in mind, we can begin by outlining first the conditions necessary for differential pricing to be emloyed successfully.

The Conditions for Differential Pricing

Three practical conditions are necessary if a seller is to practice price discrimination effectively: (1) multiple demand elasticities, (2) market segmentation, and (3) market sealing.

Multiple Demand Elasticities. There must be differences in demand elasticity among buyers due to differences in income, location, available alternatives, tastes, or other factors. If the underlying conditions that normally determine demand elasticity are the same for all purchasers, the separate demand elasticities for each buyer or group of buyers will be approximately equal and a single rather than multiple price structure may be warranted.

Market Segmentation. The seller must be able to partition (segment) the total market by segregating buyers into groups or submarkets according to elasticity. Profits can then be enhanced by charging a different price in each submarket.

Market Sealing. The seller must be able to prevent—or natural circumstances must exist which will prevent—any significant resale of goods from the lower- to the higher-priced submarket. Any leakage in the form of resale by buyers between submarkets will, beyond minimum critical levels, tend to neutralize the effect of differential prices and narrow the effective price structure to where it approaches that of a single price to all buyers.

Kinds of Differentials

In view of the above, what practical techniques can sellers use in establishing a structure of price differentials? Actually, it may be accomplished in several ways. In the following paragraphs the criteria employed will be differential structures based on: (1) quantity, (2) geographic location, (3) time, and (4) product use. This classification, it will be seen, cuts across the three degrees (forms) of price discrimination but places major emphasis on the most interesting and important one—price discrimination of the third degree.

Quantity Differentials. Three types of quantity discounts are particularly worth noting because of their significance in business practice from the standpoint of managerial pricing decisions, and for their legal aspects in government antitrust policy (treated briefly here and again in the next chapter). They are: (1) cumulative discounts, (2) quantity discounts, and (3) functional discounts. Each of these deserves some separate comment.

Cumulative Discounts. Cumulative discounts are based upon total quantity bought over a period of time (such as a year). They are granted by sellers primarily as a concession to large buyers, or for the purpose of encouraging greater buyer loyalty, or because they may reduce costs by facilitating forward planning in production, stabilize seasonal output variations, and reduce investment in inventories. In any event, a cumulative discount is worthwhile to the seller if he realizes a saving from sales made to a particular buyer over a period of time, where such savings were not reflected in the price paid by the buyer but which are reserved and refunded to him at the end of a period of time. As for their legality, however, the Federal Trade Commission observed in the H. C. Brill case (26 FTC 666, 1938) that any system of discounts based on the amount of annual sales is a price discrimination in violation of the Clayton Act (Section 2(a), as amended) if it tends substantially to lessen competition, unless justified by due allowance for differences in cost not previously allowed and resulting from quantities sold. In other words, cumulative discounts by themselves *may* not be illegal if a cost saving can be shown in the firm's accounting records, and if the discounts are proportional to the saving.

Quantity Discounts. Based upon the amount of the purchase at one time and its delivery to one location, quantity discounts are thus determined by the size of a single purchase and are granted in order to encourage larger orders so as to reduce the costs of selling, accounting, packing, delivery, etc. As for their legality, there may again be a moot issue involved. The FTC has at times found quantity discounts permissible if justified by differences in costs, e.g., Kraft-Phenix Cheese case (25 FTC 537, 1937) and American Optical Co. case (28 FTC 169, 1939). But in the circumstances of the Morton Salt case, the FTC found that the company's carload as well as cumulative discounts were not justified by differences in cost and hence were injurious to competition. The firm was ordered to desist from selling to retailers at prices lower than those charged wholesalers whose customers compete with them, and the order was later sustained by a Supreme Court decision (334 U.S. 47, 1948).

Functional Discounts. Based upon the trade classification of the buyer (e.g., wholesaler, jobber, retailer, etc.), functional discounts are also commonly referred to as "distributor discounts." Since these discounts are granted to distributors according to the latter's position in the product's channel of distribution, the various differentials have the purpose of inducing distributors to perform their particular marketing functions. From a legal standpoint, it has been held that when buyers are in competition with one another, as when dealers are at the same level in the distributive structure, discrimination practices between them ("horizontal discounts") are in violation of the law. But differences in discounts at different levels in the structure ("vertical discounts") have been held to be legal and, accordingly, the FTC has never issued an order against such discounts (in re: Standard Brands, 30 FTC 1117, 1940; Caradine Hat Co., 39 FTC 86, 1944). From the seller's standpoint, therefore, the following problems must be considered in setting an appropriate structure of differentials:

1. Buyers must be classified, not on an arbitrary basis, which would be illegal, but according to their strict nature of operations or functions undertaken. Buyers at the same level, such as mail-order houses, chain stores, and independent retailers, must be placed in the same class and then the discounts granted must not exceed cost savings (in re: Pittsburgh Plate Glass Co., 25 FTC 1228, 1937; American Oil Co., 29 FTC 857, 1939; Sherwin-Williams, 36 FTC 25, 1943).

2. Properly classified, a retailer might thus be given a larger discount than a wholesaler. But where the buyer performs more than one function, as selling both at retail and wholesale, the FTC has ruled that the larger discount can be applied only to the portion of the order for which that function alone is performed. In practice, however, the rule is difficult to enforce because the seller must take the buyer's word as to how different portions of the order will be handled, and the buyer may tend to overstate the quantity on which the higher discount applies (in re: *Southgate Brokerage Co.* vs. *FTC*, 150 F. 2d 607, 1945; *Standard Oil Co. of Indiana* vs. *FTC*, 173 F. 2d 210, 1949). But the rule is also open to criticism when applied to the dual-function dealer, since he is denied a rightful discount for performing the wholesale function on that part of the order which he retails himself.

3. Discounts must provide adequate margins to cover operating costs and normal profits of dealers. Excessive margins will encourage entry of new distributors while deficient ones will result in lost orders. Since this entails the practical difficulties of knowing dealers' costs, two useful guides to the seller for judging those costs are: (*a*) the cost of selling through alternative channels (including the alternative of bypassing dealers and performing the function himself), which sets an upper limit to the size of the discount in each case; and (*b*) the extent to which price cutting prevails at successive stages in the channel, with large cuts indicating that margins are too high, or perhaps that the least efficient dealers should be dumped and that a discount structure for the most efficient ones should be devised.

4. Industry tradition and competitive practices with respect to discounts are further factors to be considered. The seller who offers an unusually high margin may or may not succeed in increasing his turnover rate, depending upon whether the dealer can push the product and upon the seller's market share. Where product differentiation is negligible, consumers are relatively indifferent except for price, and the dealer's salesmanship may have little effect despite his greater incentive to sell. Yet, if he passes the greater margin along to the consumer by cutting price, the result may be a price war unless the seller's market share is small enough so that larger competitors do not deem it necessary to meet the lower price. Sugar refineries, gasoline stations, and several other oligopolistic industries provide many actual examples of this condition.

The above comments as to quantity differentials serve once again to emphasize the dangers inherent in using ordinary cost accounting data as a basis for management decisions. The ambiguous nature of such data must be recognized by any management interested in establishing a differential price structure, since the burden of proof of cost differences is on the seller and not on the government. The pattern that seems to be emerging from recent cases reveals a tendency of the courts to place increasing reliance on the FTC's interpretation of the situation. Management, therefore, must consider the legal as well as the economic aspects of differential pricing if such pricing is to become a basis for future policy. As stated earlier, however, the present chapter is concerned primarily with the economics of the problem, while the legal implications are treated further in the following chapter.

Geographic Differentials. Unlike quantity differentials, which attempt to exploit differences in quantity purchased, geographic price differentials can be used by sellers to exploit the differences in buyer locations. Before considering this type of pricing policy, some preliminary definitions are in order.

A seller may quote prices either at the point of origin of his goods or at their point of destination. Point-of-origin prices are more commonly known as "f.o.b. shipping point" prices, the idea being that the seller agrees to deliver the goods without charge, i.e., to place them "free on board" the conveyance provided by the buyer or the nearest common carrier (such as the nearest dock, railway station or airport). Point-of-destination prices, called "delivered prices," include the cost of shipping the goods to the buyer's location or to the common carrier (e.g., dock, station, or airport) nearest him. It is commonly believed, by many businessmen and even by some lawyers and economists, that delivered prices are discriminatory and f.o.b. prices are not. The facts are that this is usually true, but not always, and that both kinds of price quotations may sometimes be discriminatory and sometimes not, depending on particular circumstances. The ultimate economic test, it will be seen later, is not the form in which the price is quoted, but a comparison of the seller's realized receipts from different sales. In practice, a seller will frequently adopt one of several alternative geographic pricing policies, depending on the nature of his product, his transportation costs, and the industry's competitive structure. The resulting price structure will vary in each case, as indicated by the following, which represent the more common geographic pricing alternatives available to most companies.

Uniform F.O.B. Mill Pricing. Under this type of pricing, the seller charges all buyers in the same trade classification the same mill price for goods of the same quality purchased in similar quantities. Two variations may be employed: (1) the buyer pays the mill price and then selects his own means of transportation and pays his own freight costs; (2) the seller quotes a delivered price which is composed of the uniform mill price plus the actual freight to the buyer.

Neither of these methods involves any price discrimination. In both instances the seller's price structure is the same, and his return on every sale, or mill net, is the same, regardless of the buyer's location. The cost to buyers will differ only according to their distance from the mill. From the seller's standpoint, the only difference in policy between the two alternatives is that, if the second is chosen so that a delivered price is being quoted, the seller retains title while the goods are in transit and hence he must be the one to file charges with the carrier in the event of loss or damage to the goods.

What economic conditions must prevail if a firm is to practice uniform f.o.b. mill pricing successfully? Some of the more essential characteristics may be noted. The ratio of the value of the good to its transportation cost must be high. That is, transport costs must be a relatively small proportion of buyer's cost, or else the sale of the good will be confined to the seller's local market. Products must be differentiated, i.e., have a low cross-elasticity of demand, so that sellers are not under pressure to meet competing prices of nearby rivals. Fixed costs must be a relatively small percentage of total costs and marginal costs must be close to average costs at average output levels, so that there is a minimum of pressure on sellers to extend themselves into distant markets in order to break even or earn a profit. Plants must be geographically distributed in close conformity with their markets so as to minimize the number of distress (excess production) and shortage (excess demand) areas.

The above conditions are clearly the opposite of what prevails in typical oligopolistic industries, and hence raise serious doubts as to whether a general system of uniform f.o.b. mill pricing could ever be established (as many economists and legislators have proposed) without wreaking havoc with the competitive structure of most American industries. Normally, oligopolistic sellers must be allowed to absorb freight charges if they are to meet the prices of rivals in distant markets. In oligopolistic industries, a forced adherence to f.o.b. pricing without freight absorption would inevitably have the effect of reducing the number of firms in each market area and increasing the average size of the firm instead. The result would be a long-run movement toward more, rather than less, monopoly in American industry. (See also the discussion below and in the next chapter.) At one time or another, f.o.b. pricing has been employed in the sale of such goods as automobiles, agricultural machinery, apparel, household furniture, standard drugs, staple foodstuffs, and textiles.

Postage-Stamp Pricing. This type of geographic pricing is defined simply as the charging of the same delivered price at all destinations in the economy irrespective of buyer location. The actual method of quoting price may take either of two forms: (1) the seller may quote the same price at every destination, in which case his price already covers his average expenditure for freight, or (2) the seller may quote a uniform f.o.b. price to all buyers, but make allowances by permitting customers to deduct their full freight charge from their bill. In either case, economic discrimination is involved, for although prices at different destinations are the same, buyers located nearer to the seller pay more for freight than those located farther away. The Supreme Court, however, has held this type of pricing, i.e., uniform delivered pricing, to be legal, even though it involves discrimination in the economic sense.

Postage-stamp pricing is most commonly employed for goods that have a high value-transport ratio, and where the product is branded and has national distribution. The seller can thus maintain a uniform resale price at all locations and can quote the price in advertising. Examples of goods that have been priced in this way include appliances, hardware, auto accessories, typewriters, cosmetics, soft drinks, and candy bars. Occasionally, capital goods, e.g., light construction equipment and machinery replacement parts, have also been priced in this manner.

Zone Pricing. Under this type of pricing, the seller divides the economy into zones or regions and charges the same delivered price within each zone, but different prices between zones sufficient to cover his average freight costs as a whole. As before, the seller can either pay the freight himself or permit the buyer to pay it and then deduct it from the invoice. In either case, the seller's average mill net is the same in every zone. However, the seller will be discriminating within each zone and along each boundary, because: (1) he will allow less freight than is paid at a farther boundary of a zone and more than is paid at a nearer one, while (2) he allows less to buyers on the nearer side of a boundary than he does to others just across the line.

What are the legal implications of zone pricing? Evidently, if the seller's price zones are the same as the freight-rate zones, then this type of pricing is the same as f.o.b. pricing and no legal difficulties are involved. However, if the price zones do not conform with the freight-rate zones, the seller's mill net is higher for nearby customers in the zone than for farther ones, thus involving economic discrimination similar to that found in postage-stamp pricing. This is the more common case in industry and, in the light of recent court decisions on basing-point pricing (discussed next), may be open to serious questions as to legality.

Generally speaking, zone pricing is preferred where the freight cost on branded goods is too high to permit their sale throughout the country at a uniform delivered price. The more significant the freight charge, the greater the number of zones and the smaller their size. Conversely, for products that have a relatively low transportation cost, zones are normally few but large. Prices quoted in advertisements with such qualifying statements (usually in small print) as "slightly higher west of the Rockies" or "west of the Mississippi" are typical examples. Zone pricing has been widely used for a tremendous variety of products including major appliances such as washing machines, refrigerators and ranges, and also transformers, elevators, paint, power cables, soap, and book matches, to mention only a few.

Basing Point Pricing. The basing point system is a method of quoting delivered prices that has been used mainly in sales by manufacturers to other producers. A basing point price consists of a factory price plus a transportation charge. The transportation charge, however, does not always correspond to the actual cost; instead the charge is usually from some designated production center known as a "basing point." Under such a system the seller may calculate his delivered price by using either single or multiple basing points.

The outstanding example of the single basing point system is "Pittsburgh Plus," employed by the steel industry and ordered discontinued by the FTC in 1924 in a landmark case against the United States Steel Corporation. Under this system, every seller, regardless of location, would quote a buyer, also regardless of location, the Pittsburgh mill price of steel plus the rail freight from Pittsburgh to the destination, irrespective of the actual origin of the shipment or its actual freight cost. Hence the term "Pittsburgh Plus." All firms in the industry tended to follow the same practice, the price leader being the U.S. Steel Corporation, and hence buyers were usually quoted the same prices by competing sellers on most steel products.

Under the multiple basing point system, two or more producing centers are designated as basing points, and every seller then quotes a delivered price equal to the mill price at the basing point nearest the buyer plus the freight cost from that point to the destination, again irrespective of the actual origin of the shipment or its actual freight cost. The price among the various basing points may be either equal or unequal. The important thing is that each seller quotes a delivered price which is a combination of the basing point price and transport costs. (In some instances the delivered price may be the lowest combination of base price at any mill plus the rail freight from that mill to the particular destination.) Two terms associated with basing point pricing are "phantom freight" and "freight absorption," each of which should be distinguished.

Phantom freight occurs when the freight charge to the buyer exceeds the freight actually paid by the seller in making the delivery. This occurs when: (1) the buyer is closer to the nonbase selling mill than he is to the basing point, or (2) when the actual delivery is made by a carrier cheaper than rail, such as truck or barge, or (3) when the nonbase mill makes a delivery in its own city and incurs no significant transportation costs, in which case the delivery charge is virtually "pure" phantom freight. Phantom freight is thus a profit to the seller.

Freight absorption occurs when the freight charge to the buyer is less than the freight cost actually incurred by the seller in making the delivery. It can occur in circumstances that are the opposite of those for phantom freight, but typically it arises when the buyer is closer to the basing point than he is to the nonbase selling mill. Clearly, therefore, only when a seller's mill is a basing point and only when he sells in an area where that basing point governs the delivered price will his delivered price equal his mill price plus freight. In this case there is no phantom freight and no freight absorption, or in effect the price is the same as an f.o.b. mill price. In practice, however, sellers do reach out into distant markets and quote the same delivered prices as competitors, thereby willingly absorbing freight which, however, is more than offset by the reduction in average costs resulting from the fuller use of capacity. Hence, they obtain variable mill-net yields.

What are the economic conditions favorable to basing point pricing, and what of its legality? There are two schools of thought on the subject. One school, composed of such economists as Frank Fetter, Fritz Machlup, and Vernon Mund, and antitrust agencies such as the FTC and the Department of Justice, holds that the basing point system is a means of assuring identical delivered prices, that it enables sellers to eliminate price competition, that it is *prima facie* evidence of collusion, and that therefore it should be generally outlawed. A second school, including J. M. Clark, Melvin de Chazeau, and a number of other noted business economists, argues that the basing point system is a unique but normal outgrowth of oligopolistic industries, that superficial changes of the pricing methods in these industries will not alter their fundamental competitive structure nor produce the pricing results of perfect competition, and therefore to outlaw it universally will only result in some equivalent, or possibly less desirable, pricing practices. Their reasons are based on the typical economic characteristics of oligopolistic industries: (1) inelastic demand, (2) standardized products, (3) relatively large proportion of fixed costs to total costs and fairly constant per-unit variable costs to near-full capacity, resulting in high break-even points, (4) heavy transportation charges, and (5) scattered plant locations in relation to markets. The steel and cement industries are often cited as typical examples, although the multiple basing point system has also been employed in the sale of lead, lumber, sugar, pulp, and a wide variety of other products, especially heavy goods.

As to its legality, the law does not appear too clear. In various cases involving cement, rigid steel conduit, and corn products producers tried since World War II, the courts have upheld the FTC's charges that the basing point system is in violation of the Clayton Act and therefore illegal. Congress, however, has refrained from universally outlawing the system and the Supreme Court has consequently taken the alternative of evaluating each case separately. Though many firms still follow this type of pricing policy, there may be serious doubt as to its legality in the light of these recent cases.

Freight-Equalization Pricing. Under freight-equalization pricing, the seller charges the buyer a freight cost which he would pay in getting delivery from a nearer supplier. The seller may accomplish this by quoting a delivered price that covers freight from his competitor's mill but pay the higher freight from his own mill, or he may quote an f.o.b. price or a delivered price covering his own freight, and allow the customer to deduct from his bill the excess freight over and above that which would be charged by a competitor closest to the buyer. In either case the seller quotes identical delivered prices of competitors by absorbing freight. The seller's return, or mill net, thus varies, depending on the amount of freight he absorbs on each sale; hence, as in previous instances, he is discriminating in price. The seller may follow this type of policy occasionally in order to utilize excess capacity. If done generally and *systematically*, however, it is likely to be illegal in the light of the recent court cases mentioned above. Typically, freight-equalization pricing has been found in industries characterized by standardized products, many sellers, high fixed costs, heavy investment in fixed assets, and a low value-transport ratio. Examples include bituminous coal, lumber, and gasoline, among others.

Time Differentials. As a third classification of differential pricing, we can consider the phenomenon of temporal discrimination. Market segmentation, instead of being achieved by exploiting differences in quantity purchases or buyer locations, as in the two previous classifications, is now accomplished through the medium of *time*. As in other kinds of price differentials, the object from the seller's standpoint is to capitalize on the fact that buyers' demand elasticities vary, but in this case as a function of time. Thus two classes of price differentials may be distinguished, extending from the narrowest to the broadest "slice of time."

Clock-Time Differentials. When demand elasticities of buyers vary within a 24-hour period, the seller has the opportunity of exploiting these differences through price differentials. The most common examples of this are the differences between day and night rates on long-distance telephone calls, and the differences between matinee and evening admission charges in movies and theaters. When price differentials are based on clock time, the object of the seller is to charge a higher price for the product in the more inelastic period and a lower price during the more elastic interval. Telephone rates and theater prices are thus an interesting contrast. With the former, the more inelastic demand period is during the day and with the latter it is during the evening; conversely, demand for long-distance phone calls is more elastic in the evening, while the demand for movies and theater is more elastic in the daytime. Prices are thus structured accordingly so as to utilize the advantages of these differences in buyers' time preferences.

What economic conditions are necessary to make the construction of clock-time price differentials profitable? Three typical factors may be noted: (1) Buyers must have a definite and strong preference for purchasing at some times rather than others. This gives rise to a significant difference in demand elasticities as a function of time. (2) The seller must have the facilities for providing the product to buyers in slack periods at prices that will cover variable costs and also contribute to the recovery of some fixed costs. Thus, considerations have at times been given by public service commissions to raise the rates on public transportation facilities (e.g., subways) during rush hours in order to increase revenues, reduce congestion, and distribute the use of facilities more evenly. In some of these cases it has been found, however, that the incremental cost and inconvenience of administration (e.g., converting or adjusting turnstiles twice in the morning and twice in the evening) would make the process unfeasible if not unprofitable. (3) The product's use must be nonstorable either wholly or in part. That is, the buyer must consume the entire product at one time and in the time interval for which he pays, or else leakage will occur as resale markets develop for the new or partly used product.

Calendar-Time Differentials. Price differentials may be based not only on elasticity differences within a day, but on differences between days, weeks, months, or seasons as well. In addition to telephone rates and theater prices, which exhibit weekend variations in addition to intraday (i.e., clock-time) price differences, other examples of calendar-time price differentials are found in the sale of services by recreational facilities such as golf courses, tennis courts, and swimming pools; the sale of food by some restaurants; and seasonal variations in the sale of clothing, resort accommodations, and vacation trips. Calendar-time differentials thus refer to any variable price structure based on time that extends beyond the 24-hour period of clock time. Seasonal variations, since they occur within a year and are due strictly to weather and custom, are more broadly a function of time in that variations in weather and in custom (e.g., Christmas and Easter buying) are recurrent and fairly periodic. Hence, seasonal variations may justifiably be placed in the category of calendar-time differentials from the standpoint of the seller who is considering this type of pricing structure. Perhaps cyclical variations could also be included if they were regular and periodic in the calendar sense, which they are not for the economy as a whole but may be for certain (relatively few) business firms. In any case, as with clock-time differentials, the object of the seller is to derive a price structure that exploits the time preferences of buyers. For many products the economic characteristics stated above with respect to time preferences, cost considerations, and the nonstorability of product use are relevant in setting calendar-time differentials well. But other than these factors, special conditions may prevail in particular circumstances that would make any general statement inapplicable.

Legality. The setting of prices according to time differentials is discriminatory when such prices are unrelated to cost differences or to differences in satisfaction provided by the product. Conversely, maintaining the same price despite variations in cost would also be a form of temporal price discrimination. Since demonstrated differences in cost as shown by the seller are the chief test as to the legality of a discriminatory pricing policy, there is the problem of calculating costs to show fluctuations over time. The accounting problem of measurement is again posed. For instance, should fixed costs be allocated among sales made at different hours, days, weeks and months, or should they be distributed equally for each dollar of sales irrespective of the time factor? The first alternative would raise sharply conflicting theoretical problems and hence measurement difficulties, while the second may miss the real target by a wide mark. As of this time, the antitrust agencies have paid relatively little attention to temporal price discrimination and, in the light of their limited budgets, it seems likely that this trend will continue. Time differentials thus offer a profitable area for research and development on the part of many business firms.

Product-Use Differentials. A fourth classification as a basis for price discrimination is to segregate buyers according to their use of the product. The classical application of product-use differentials dates back to the nineteenth century in the long- and short-haul pricing practices of railroads. These amounted simply to charging what the traffic would bear by sometimes pricing a short haul higher than a long haul over the same line under substantially similar circumstances, according to competition at various points.

In product-use discrimination, the problem of the seller is to carve the market up into homogeneous groups according to demand elasticity as determined by the buyer's use of the product. A variety of examples may be cited as illustrations. In the service industries, electric and gas companies establish separate rate structures for residential and commercial users; telephone companies distinguish between residential and business phones; movie theaters, barber shops, and public carriers set separate charges for adults and children despite equal time and space costs of serving both groups; and railroads sell freight transportation service at different prices to different groups according to the goods shipped. In manufacturing, the glass container industry sold identical containers as domestic fruit jars and as packer's ware, the former at substantially higher prices because of a much lower demand elasticity; Du Pont and Rohm and Haas sold methyl methacrylate for commercial purposes at 85 cents per pound, but for denture purposes it sold to the dental profession at $45 per pound; the Aluminum Company of America used to sell aluminum ingots at a higher price per pound than it sold aluminum in cable form (on the condition, in the latter group, that the buyer would not melt it); and similarly, plate glass manufacturers sold their product at a substantially higher price per square foot for large pieces than for small pieces even though all plate glass is produced in large sheets, the reason being that competition in the small-piece market was much more severe due to the competition of ordinary window glass. In agriculture, it is well known that, usually, the price paid farmers for milk depends on the use to be made of the milk, whether for bottling or for manufacture into butter, cheese, or ice cream.

What conditions must exist for a seller to establish an effective structure of price differentials according to product use? At least two conditions are essential: first, there must be a difference in demand elasticity among buyers as to product use; and second, the seller must be able to segment these buyers into fairly homogeneous groups. The second condition, which is one of implementation, is typically accomplished by differentiating products as to design, quality, brand name, time of sale, or distribution channel, each having a different appeal to different customers. These techniques are also commonly employed as market segmentation devices in the other forms of differential pricing discussed previously. Grade labeling, prestige advertising, and similar tactics commonly encountered are merely modifications of these techniques and are often used for the purpose of sealing markets as a means of price discrimination.

Legality. In principle, as always, a chief antitrust test of legality in price discrimination is whether differences in prices are warranted by differences in costs. In practice, it is impossible to state categorically whether product-use differentials are legal or not. The procedure of the courts is to consider each case separately. Nor are there any clear-cut trends toward illegality emerging as have appeared, for example, in basing point pricing. The difficulties of measuring cost, the widespread application of product-use discrimination in industry, and the obscure nature of the law on the subject of discrimination in general—all these are factors to consider if management is to establish an appropriate policy of differential pricing. Other than this, any further comments as to the legality of price discrimination are reserved for the following summary and the next chapter.

Summary and Conclusions

Differential pricing is a practical, multidimensional technique available to management as a means of enlarging profits. It exploits the differences in demand elasticities among buyers as a basis for establishing prices. Although there are different ways in which this exploitation can be accomplished, four common ones include quantity differentials, locational differentials, time differentials, and product-use differentials. In many industries it is possible to employ these approaches in combination as well as separately.

Market segmentation and market sealing are necessary if differential pricing is to be effective. Several techniques for accomplishing segmentation and sealing are available, including variations in product design, quality, branding, choice of channel, time of sale, conditions of sale, patents, packaging, and advertising. Each of these offers opportunities for dividing the market and increasing revenues, and hence represents a vast area for management research and experimentation.

The legal aspects of differential pricing are not at all clear in many respects. It is beyond the scope of this work to go into the details of antitrust economics. Perhaps the only general statement that can be made is that all systems of pricing other than f.o.b. mill pricing may be subject to legal attack by governmental agencies such as the FTC and the Department of Justice, particularly when there are implications of collusion, conspiracy, and attempts to monopolize. Since this is not the case in many instances, various forms of differential pricing are widespread and probably will continue to exist for a long time to come. Businessmen, therefore, can find in differential pricing an opportunity for enlarging profits and facilitating their company's growth, provided such price structures are established with the guidance of sound economic principles and competent legal counsel.

BIBLIOGRAPHICAL NOTE

A significant contribution toward integrating marketing price practices with the general theory of pricing is E. R. Hawkins, "Price Policies and Theory," *Journal of Marketing* (January, 1954). Cost-plus pricing, with particular reference to British experience, is treated in the well-known study by R. Hall and C. Hitch, "Price Theory and Business Behavior," *Oxford Economic Papers No. 2* (May, 1939). This study aroused substantial interest in the theory of full-cost pricing and later resulted in the famous Lester-Machlup controversy. See R. Lester, "Shortcomings of the Marginal Analyses for Wage-Employment Problems," *American Economic Review* (1946); F. Machlup, "Marginal Analysis and Empirical Research," *ibid.*, and the respective replies and rejoinders in the 1947 volume along with the article by Stigler, "Professor Lester and the Marginalists." Other works analyzing businessmen's views and practices include J. Earley, "Cost Accounting and the 'Marginal Analysis,'" *Journal of Political Economy* (1955), the same author's "Marginal Policies of 'Excellently Managed' Companies," *American Economic Review* (1956), and C. Saxton, *The Economics of Price Determination.* On product-line (multiple-product) pricing, three useful sources are Stigler, *Theory of Price*, 1st ed., chap. 16, Weintraub, *Price Theory*, chap. 14, and Dean, "Problems of Product-Line Pricing," *Journal of Marketing* (January, 1950). The loss-leader problem appears more fully in C. Roos, *Dynamic Economics*, and the two-part tariff in an article of that name by W. Lewis, *Economica* (1941). Finally, on differential pricing (or price discrimination), some pathbreaking general works include J. Robinson, *Economics of Imperfect Competition*, chaps. 15 and 16, and A. C. Pigou, *Economics of Welfare*, chap. XVII.

Readers desiring a more general survey of the subjects treated in this chapter will find adequate material in Dean, chaps. 7 through 9; Howard, chap. XII; A. Oxenfeldt, *Industrial Pricing and Market Practices*, chaps. 4 and 5; F. Machlup, *The Political Economy of Monopoly*, chap. 5; S. Nelson and W. Keim, "Price Policy and Business Behavior," *TNEC Monograph No. 1;* and C. Wilcox, *Public Policies Toward Business*, chaps. 7 and 8, as well as most other works on public control which place greater stress on the legal aspects of pricing. An excellent collection of short readings, skillfully edited and many of them leading contributions, is available in J. Backman, *Price Practices and Price Policies*. Chapters 5 and 12 are especially suitable as supplements to this chapter.

QUESTIONS

1. What assumptions underlie the use of odd pricing? Illustrate graphically.
2. How might "psychological pricing" be distinguished graphically from odd pricing?
3. In general, what conditions are necessary for the type of demand curve of Figure 8–4, p. 283, to prevail?
4. Examine Figure 8–5, p. 285, and discuss the economic significance of the portion of the curve below the bend.
5. (*a*) Explain the meaning of price lining. (*b*) Does price lining make for greater or less price flexibility? Explain. (*c*) Can you suggest an alternative pricing method that would likely be more profitable? Defend your answer.
6. (*a*) Explain the logic behind the use of quantity discounts. (*b*) Why do utility companies typically maintain separate discount schedules for commercial and household users?
7. What advantages are there to expressing pricing practices in a theoretical vein as done in this chapter?
8. Evaluate in your own words the method of cost-plus pricing by (*a*) stating briefly its nature, and (*b*) its pros and cons.
9. (*a*) How does flexible-markup pricing differ, if at all, from cost-plus pricing? (*b*) Is the former widely used by businessmen? Explain.
10. What do you believe to be the most fundamental difference between intuitive pricing and experimental pricing?
11. Outline some of the factors which, in practice, tend to make for stable price structures on the part of most manufacturers.
12. (*a*) Explain the meaning of price leadership. (*b*) What economic conditions are typically needed for effective price leadership? (*c*) What sort of advantages may accrue to price followers?
13. Evaluate the "fair" profit motive as a justification of cost-plus or other outdated pricing methods of management.
14. What is the basic objective in establishing a product-line pricing policy?

15. (*a*) Distinguish between loss leaders, tie-in sales, and the two-part tariff. (*b*) What do these have in common? (*c*) Discuss the economics of each.
16. Formulate a definition of differential pricing. How does it relate to price discrimination?
17. (*a*) Distinguish between the various "degrees" of differential pricing. (*b*) What conditions are needed for effective differential pricing? Explain.
18. Outline the types of differential price structures commonly used in industry. Explain each.

Chapter 9 COMPETITION AND CONTROL

The American economy is essentially a competitive economy, and as such most decisions that are made by managers are essentially competitively oriented. It is appropriate, therefore, to include in a book that is primarily concerned with management decision making a chapter on the nature of competition and its framework of legal controls. For management decision making does not take place in an economic vacuum, but rather in a sociopolitical environment that must be recognized as a limiting factor in the process of adjusting to uncertainty. Decisions and plans, in other words, may sometimes have to be modified from what they otherwise would have been if economic principles were the sole guide for action.

It is beyond the scope of this book to delve into the wide variety of areas that would normally be included in a full-scale study of competition and control. Instead we shall examine a few of the more important topics that are of particular interest to manufacturers from the standpoint of market economics and marketing policy. These include the nature of the antitrust laws, which are the basis for the discussion, and the relation of these laws to competitive practices in the areas of patent and trade-mark policy, exclusion and discrimination, delivered pricing, distribution, and the measurement of monopoly power. Thus, the protection and regulation of agriculture, labor, investors, utilities, and so forth, which are discussed in most books dealing with government and business, will not be treated here.

THE ANTITRUST LAWS

What are the antitrust laws? They are a number of acts passed by Congress since 1890 by which the United States government is committed to prevent monopoly and to maintain competition in American industry. Although there are also state antitrust laws in almost every state in the country, these are largely ineffectual and spasmodically enforced, since they are powerless to control agreements or combinations in major industries whose activities extend into interstate commerce. This, coupled with inadequate funds, has left the task of maintaining competition via antitrust law enforcement almost entirely to the federal government. Thus it is the federal antitrust laws that will be of concern to us here. These laws include the Sherman Act, the Clayton Act, the Federal Trade Commission Act, the Robinson-Patman Act, the Wheeler-Lea Act, and the Celler Antimerger Act. There are also others, but they are of relatively lesser significance.

Provisions of the Laws

The substantive provisions of the antitrust laws may be outlined as follows.

The Sherman Act (1890). This was the first attempt by the federal government to regulate the growth of monopoly in the United States. The provisions of the law were concise (probably too concise) and to the point. It forbade: (1) every contract, combination, or conspiracy in restraint of trade which occurs in interstate or foreign commerce, and (2) any monopolization or attempt to monopolize, or conspiracy with others in an attempt to monopolize, any portion of trade in interstate or foreign commerce. Violations of the Act were made punishable by fines and/or imprisonment and persons injured by violators could sue for triple damages.

The Act was surrounded by a cloud of uncertainty by failing to state precisely which kinds of actions were prohibited. Also, no special agency existed to enforce the law until 1903, when the Antitrust Division of the Department of Justice was established under an Assistant Attorney General. In order to put some teeth into the Sherman Act, therefore, Congress passed the Clayton Act and the Federal Trade Commission Act.

The Clayton and Federal Trade Commission Acts (1914). Aimed at practices of unfair competition, the Clayton Act was concerned with four specific areas: price discrimination, exclusive and tying contracts, intercorporate stockholdings, and interlocking directorates. About each of these it had this to say:

1. For sellers to discriminate in prices between purchasers of commodities is illegal. However, such discrimination is permissible where there are differences in the grade, quality, or quantity of the commodity sold; where the lower prices make due allowances for cost differences in selling or transportation; and where the lower prices are offered in good faith to meet competition. Illegality exists where the effect is substantially to lessen competition or tend to create a monopoly.

2. For sellers to lease, sell, or contract for the sale of commodities on condition that the lessee or purchaser not use or deal in the commodity of a competitor is illegal, if such exclusive or tying contracts substantially lessen competition or tend to create a monopoly.

3. For corporations engaged in commerce to acquire the shares of a competing corporation, or the stocks of two or more corporations compet-

ing with each other, is illegal if such intercorporate stockholdings substantially lessen competition or tend to create a monopoly.

4. For corporations engaged in commerce to have the same individual on two or more boards of directors is an interlocking directorate, and such directorships are illegal if the corporations are competitive and if any one has capital, surplus, and undivided profits in excess of $1 million.

Thus, price discrimination, exclusive and tying contracts, and intercorporate stockholdings were not declared by the Clayton Act to be absolutely illegal, but rather, in the words of the law, only when their effects "may be to substantially lessen competition or tend to create a monopoly." On interlocking directorates, however, the law made no such qualification: the fact of the interlock itself is illegal, and the government need not find that the arrangement results in a reduction in competition.

The Federal Trade Commission Act in these respects served primarily as a general supplement to the Clayton Act by stating broadly and simply that "unfair methods of competition in commerce are hereby declared unlawful." But what significant contribution to monopoly control was made by these laws? Essentially, both the Clayton Act and Trade Commission Act were directed toward the prevention of abuses, whereas the Sherman Act emphasized the punishment of abusers. To be sure, the practices that were prohibited in the two later laws could well have been attacked under the Sherman Act as conspiracies in restraint of trade or as attempts to monopolize, but now the nature of the problem was brought more sharply into focus. Moreover, under the Federal Trade Commission Act, the FTC was established as a governmental antitrust agency with federal funds appropriated to it for the purpose of attacking unfair competitive practices. No longer was it necessary to await private suits brought by private parties on their own initiative and expense in order to curb unfair practices in commerce.

In addition to the heart of the Federal Trade Commission Act, which makes unfair methods of competition illegal as quoted above, the FTC is also authorized under the Act to safeguard the public by preventing the dissemination of false and misleading advertisements with respect to foods, drugs, cosmetics, and therapeutic devices used in the diagnosis, prevention, or treatment of disease. In this respect it supplements in many ways the activities of the Food and Drug Administration which, under the Food, Drug, and Cosmetic Act (1938) outlaws adulteration and misbranding of foods, drugs, devices, and cosmetics moving in interstate commerce.

The Robinson-Patman Act (1936). Frequently referred to as the "Chain Store Act," the Robinson-Patman Act was passed for the purpose of providing economic protection to independent retailers and wholesalers such as grocers and druggists, from "unfair discriminations" by large sellers attained "because of their tremendous purchasing power." The law was an outgrowth of the increasing competition faced by independents that came with the development of chain stores and mass distributors after

World War I. Those who favored the bill contended that the lower prices charged by these large organizations were attributable only in part to their lower costs, and moreso if not entirely to the sheer weight of their bargaining power which enabled them to obtain unfair and unjustified concessions from their suppliers. The Act was thus a response to the cries of independents who demanded that the freedom of suppliers to discriminate be more strictly limited.

The Act, which amended Section 2 of the Clayton Act relating to price discrimination, contained the following essential provisions:

1. The payment of brokerage fees where no independent broker is employed is illegal. This was intended to eliminate the practice of some chains of demanding the regular brokerage fee as a discount when they purchased direct from manufacturers. The argument posed was that such chains obtained the discount by their sheer bargaining power and thereby gained an unfair advantage over smaller independents that had to use and pay for brokerage services.
2. The making of concessions by sellers, such as manufacturers, to buyers, such as wholesalers and retailers, is illegal unless such concessions are made to all buyers on proportionally equal terms. This provision was aimed at preventing advertising and promotional allowances from being granted to large-scale buyers without allowances being made to competing buyers on proportionally equal terms.
3. Other forms of discrimination, such as quantity discounts, are illegal where they substantially lessen competition or tend to create a monopoly, either among sellers or among buyers. However, price discrimination is not illegal if the differences in prices make "due allowances" for differences in cost or if offered "in good faith to meet an equally low price of a competitor." But even where discounts can be justified by lower costs, the FTC is empowered to fix quantity limits beyond which discounts may not be granted, if it believes that such discounts would be "unjustly discriminatory or promotive of monopoly in any line of commerce."
4. It is illegal to give or to receive a larger discount than that made available to competitors purchasing the same goods in equal quantities. Also, it is illegal to charge lower prices in one locality than in another for the same goods, or to sell at "unreasonably low prices," where either of these practices is aimed at "destroying competition or eliminating a competitor."

The Wheeler-Lea Act (1938). An amendment to part of the Federal Trade Commission Act, the Wheeler-Lea Act was passed for the purpose of providing consumers, rather than just business competitors, with protection against unfair practices. The Act makes illegal "unfair or deceptive acts or practices" in interstate commerce. Thus, a consumer who may be injured by an unfair trade practice is, before the law, of equal concern with the merchant who may be injured by an unfair competitive practice. The Act also defines "false advertising" as "an advertisement other than labeling which is misleading in a material respect," and makes the definition applicable to advertisements of foods, drugs, curative devices, and cosmetics.

The Celler Antimerger Act (1950). The Celler Antimerger Act is an extension of Section 7 of the Clayton Act relating to intercorporate stockholdings. The latter law, as stated earlier, made it illegal for corporations to acquire the stock of competing corporations. But that law, the FTC argued, left a loophole through which monopolistic mergers could be effected by a corporation acquiring the *assets* of a competing corporation, or by first acquiring the stock and, by voting or granting of proxies, acquiring the assets. Moreover, the Supreme Court in several cases held that such mergers were not illegal under the Clayton Act if a corporation used its stock purchases to acquire the assets before the FTC's complaint was issued (in re: *FTC* vs. *Western Meat Co., Thatcher Mfg. Co, and Swift and Co.*, 272 U.S. 554, [1926]) or before the Commission had issued its final order banning the stock acquisition (in re: *Arrow-Hart and Hegeman Electric Co.* vs. *FTC*, 291 U.S. 587 [1934]).

The Antimerger Act plugged the loophole in the Clayton Act by making it illegal for a corporation to acquire the stock *or assets* of a competing corporation where the effect may be "substantially to lessen competition, or to tend to create a monopoly." The Act thus bans all types of mergers—horizontal (similar plants under one ownership, such as steel mills), vertical (dissimilar plants in various stages of production, integrated under one ownership), and conglomerate or circular (dissimilar plants and unrelated product lines)—provided the Commission can show that the effects *may* substantially lessen competition or tend towards monopoly. It should be noted, however, that the intent of Congress in passing the Act was that there be a maintenance of competition. Accordingly, the Act was intended to apply to mergers with large firms or large with small firms, but not to mergers among small firms which may be undertaken to strengthen their competitive position.

Enforcement of the Laws

Before concluding this section, a few comments should be made as to the enforcement of the antitrust laws, since some knowledge of this is important to businessmen in carrying out activities which, they may later be surprised to learn, are the subject of a governmental investigation. Accordingly, the following paragraphs provide a brief summary of the nature and scope of enforcement as it exists at the present time.

In general, the application of the antitrust laws is effected on a case-by-case basis. That is, an order or decision resulting from an action is not applicable to all of industry, but only to the defendants in the particular case. Cases tried under the Sherman Act may originate in the complaints of injured businessmen, suggestions made by other government agencies, or in the research of the Antitrust Division of the Department of Justice, since it is this organization which may bring into the federal courts criminal or civil suits against violators of the Act. About 90 per cent of the cases, it has been estimated, arise from complaints issued by injured parties, and at the present time most of the ensuing investigations are conducted by the FBI. The Federal Trade Commission Act, on the other hand, is enforced by the FTC and, when their orders become final, through suits brought by the Department of Justice. Finally, with respect to the Clayton Act, both the FTC and the Justice Department have concurrent jurisdiction in its enforcement, and in practice it is usually a matter of which agency gets there first.

Sherman and Clayton Acts. Section 14 of the Clayton Act fixes the responsibility for the behavior of a corporation on its officers and directors and makes them subject to the penalties of fine or imprisonment for violating the laws. Under the Sherman Act, the fine is limited to $5,000, but fines have actually been pyramided into a hundred thousand dollars and more in a single case by exacting the $5,000 on each count of an indictment (e.g., monopolizing, attempting to monopolize, conspiring, and restraining trade) and by imposing the fine on each of the defendants in a suit (e.g., a trade association, each member of the association, and each of the directors and officers of the member firms). Other penalties are also possible as provided in other acts.

Businessmen who want to avoid risking violation of the law may consult with the Justice Department by presenting their proposed plans for combination or other particular practices. If the plans appear to be legal, the Department may commit itself not to institute future criminal proceedings, but it will reserve the right to institute civil action if competition is later restrained. The purpose of a civil suit is not to punish, but to restore competition by providing remedies. Typically, three classes of remedies are employed:

1. Dissolution, divestiture, and divorcement provisions may be used. Examples include an order to dissolve a trade association or combination, to sell intercorporate stockholdings, or to dispose of ownership in other assets. The purpose of these actions is to break up a monopolistic organization into smaller but more competitors.

2. An injunction may be issued. This is a court order requiring that the defendant refrain from certain business practices, or perhaps take a particular action that will increase rather than reduce competition.

3. A consent decree may be employed. This is usually worked out between the defendant and the Justice Department without a court trial. The defendant in this instance does not declare himself guilty, but agrees nevertheless to abide by the rules of business behavior set down in the decree. This device is now one of the chief instruments employed in the enforcement of the Sherman and Clayton Acts.

Finally, the laws are also enforced through private suits. Under the

Sherman Act, injured parties (individuals, corporations, or states) may sue for treble damages including court costs, and under the Clayton Act, a private plaintiff may also sue for an injunction—a restraining order—whenever he is threatened by loss or damage resulting from some firm's violation of the antitrust laws.

Federal Trade Commission Act. Under this law, the FTC is authorized to prevent unfair business practices as well as to exercise, concurrently with the Justice Department, enforcement of the prohibited provisions of the Clayton Act as amended by the Robinson-Patman Act. Accordingly, the FTC has taken action against agreements that have tended to curtail output, fix prices, and divide markets among firms, thereby striving to maintain competition as well as to prevent unfair methods.

In enforcing the laws relating to monopoly, unfair trade, and deception (including such laws as the Export Trade Act, the Wool Products Labeling Act, the Fur Products Labeling Act, and the Flammable Fabrics Act), the FTC utilizes three procedures: (1) the cooperative method, which involves conferences on an individual and industry-wide basis in order to secure voluntary compliance by businessmen with respect to the rules of fair competition; (2) the consent method, whereby the Commission may issue a stipulation to the violator stating that he discontinue his illegal practices; and (3) the compulsory method, which involves legal action based upon the issuance of formal complaints. In general, the Commission obtains its evidence for making complaints from its own investigations, from injured competitors, from consumers, and from other governmental agencies. About 10 per cent of the cases actually selected arise from the Commission's own investigations; the remaining 90 per cent are derived from the other sources, particularly from the complaints of injured parties.

Summary

The chief prohibitions contained in the antitrust laws, together with the relevant sections and acts, may now be summarized as follows:

1. It is flatly *illegal*, without any qualification, to:
   a) enter a contract, combination, or conspiracy in restraint of trade (Sherman Act, Sec. 1);
   b) monopolize, attempt to monopolize, or combine or conspire to monopolize trade (Sherman Act, Sec. 3).
2. When and if the effect may be substantially to lessen competition or tend to create a monopoly, it is *illegal* to:
   a) acquire the stock of competing corporations (Clayton Act, Sec. 7);
   b) acquire the assets of competing corporations (Clayton Act, Sec. 7, as amended by the Antimerger Act in 1950);
   c) enter exclusive and tying contracts (Clayton Act Sec. 3);
   d) discriminate unjustifiably among purchasers (Clayton Act, Sec. 2, as amended by Robinson-Patman Act, Sec. 1).

3. In general, it is also *illegal* to:
   a) engage in particular forms of price discrimination (Robinson-Patman Act, Sec. 1 and 3);
   b) serve as a director of competing corporations of a certain minimum size (Clayton Act, Sec. 8);
   c) use unfair methods of competition (Federal Trade Commission Act, Sec. 5);
   d) use unfair or deceptive acts or practices (Federal Trade Commission Act, Sec. 5, as amended by Wheeler-Lea Act, Sec. 3).

Thus the laws taken as a whole are designed not only to prevent the growth of monopoly, but to maintain competition as well. The extent to which they have succeeded in accomplishing these ends forms the chief purpose of the discussion in the following section.

AREAS OF UNCERTAINTY

The antitrust laws have various things to say about monopoly, competition, and related concepts. Specifically, the Sherman Act forbade restraints of trade, monopoly, and attempts to monopolize; the Clayton Act forbade certain practices where the effects may be to lessen substantially the degree of competition or tend to create a monopoly; and the Federal Trade Commission Act forbade unfair methods of competition. But although Congress succeeded in passing these laws, it failed to define, and left up to the courts to interpret in their own way, the meaning of such terms as "monopoly," "restraint of trade," "substantial lessening of competition," "unfair competition," and so on. From the management standpoint, therefore, these create areas of uncertainty that need to be understood if decisions are to be made and plans formulated that will guide the firm's future course of action.

But how are these areas to be understood, and in what connection? The most suitable method is to approach the problem from the standpoint of particular issues they raise. Accordingly, the following paragraphs will consider a number of problems that are of particular concern to the executive in the field of competition and control. Since judicial interpretation has been crucial in determining the applications and effects of the antitrust laws, we shall attempt to sketch briefly the nature of each issue, some leading court decisions, and the major trends. In this way, it will become evident that there are degrees of uncertainty both within and between issues, a knowledge of which provides management with a sounder base on which to plan future policies.

Restrictive Agreements

The state of the law as to restrictive agreements of virtually any type among competitors is reasonably clear, and the courts have almost always, with few minor exceptions, upheld the government in such cases. In general, a restrictive agreement is regarded by the government as one that results in a restraint of trade among separate companies. It is usually understood to involve a direct or indirect, overt or implied, form of price fixing, output control, market sharing, or exclusion of competitors by boycotts or other coercive practices. It makes no difference whether the agreement was accomplished through a formal organization such as a trade association, informally, or even by habitual identity of behavior frequently referred to as "conscious parallel action" (e.g., identical price behavior among competitors). It is the effect, more than the means, that is judged. Thus in the second American Tobacco case in 1946, it was charged that the "big three" cigarette producers exhibited striking uniformity in their buying prices on tobacco and in their selling prices on cigarettes, as well as in other practices. Despite the fact that not a shred of evidence was produced to indicate that a common plan had even so much as been proposed, the Court declared that conspiracy "may be found in a course of dealings or other circumstances as well as in an exchange of words," and hence the companies were held in violation of the law (328 U.S. 781, 810). In other words, no secret meetings in a smoke-filled room and no signatures in blood are needed to prove the conspiracy provisions of the Sherman Act. Any type of agreement, explicit or implicit, any practice, direct or indirect, or even any action with the knowledge that others will act likewise to their mutual self-interest, is likely to be interpreted as illegal if it results in exclusion of competitors from the market, restriction of output or of purchases, division of markets, price fixing, elimination of the opportunity or incentive to compete, or coercion. To some extent, the doctrine of conscious parallel action by which firms can be convicted on rather flimsy circumstantial evidence has, fortunately, been partially repudiated by judges in more recent cases. However, it still remains as a fairly significant antitrust barometer and is likely to be used, though perhaps more sparingly, in the foreseeable future.

Combination and Monopoly

Concerning monopoly, the state of the law is less certain and the position of the courts less consistent than in cases involving restrictive agreements. There are three aspects of the problem to be considered: monopoly *per se;* vertical integration; and mergers.

Monopoly by Itself. Here there has been a fundamental change in the attitude of the courts since 1945. Prior to that time, it was the position of the Court that the mere size of a corporation, no matter how impressive, is no offense, and that it requires the actual exertion of monopoly power, as shown by unfair practices, in order to be held in violation of the law. But the decisions handed down in various antitrust cases since 1945 have reversed this outlook almost completely. In the case against the Aluminum Company of America in 1945, in which Judge Learned Hand turned the trend in judicial thinking on monopoly (148 F 2d 416), it was the court's opinion that: (1) to gain monopolistic power even by growing with the market, i.e., by reinvesting earnings rather than by combining with others, is nevertheless illegal (p. 431); (2) the mere size of a firm *is* an offense, for the power to abuse and the abuse of power are inextricably intertwined (pp. 427–28); (3) the Company's market share was 90 per cent and that "is enough to constitute a monopoly; it is doubtful whether 60 or 64 per cent would be enough; and certainly 33 per cent is not" (p. 424); and (4) the good behavior of the Company which, prior to 1945, would have been an acceptable defense to the Court, is no longer valid, for "Congress did not condone 'good' trusts and condemn 'bad' ones; it forbade all" (p. 427).

With this decision, Judge Hand put an end to the "good-trust-vs.-bad-trust" criterion that had been used by the courts for almost a quarter of a century, beginning with the U.S. Steel case in 1920 and supplemented by the International Harvester case in 1927. And despite the doubtfulness of the measure of monopoly power and hence whether the charge of monopoly was really proven in this case, subsequent court decisions have never repudiated the doctrines enunciated by Judge Hand, although they have tempered them somewhat. Thus, at the present time, the judgment of monopoly is based on such factors as the number and strength of the firms in the market, their effective size from the standpoint of technological development and competition with substitutes and with foreign trade, national security interests in maintaining strong productive facilities and maximum scientific research, and the public's interest in lower costs and uninterrupted production (as later stated in 1950 by Judge Knox in his decree for a remedy in the aluminum case [91 F. Supp. 333, 347]). The trend, on the basis of recent cases, indicates that monopoly may be held illegal without requiring proof of intent and even if the power were lawfully acquired; and the power may be condemned even if never abused, especially if it tends to limit or bar market access to other firms.

Vertical Integration. Here the Court stated, in the Paramount Pictures case in 1948, that such integration might be illegal if it were undertaken "to gain control over an appreciable segment of the market and to restrain or suppress competition," or if there was evidence of a power and intent to exclude competitors (334 U.S. 131, 174). But integration *per se*, it said, was not illegal. The Paramount case, which was one of the most important disintegration cases in recent years, involved five major motion picture producers operating first-run theaters in large cities and a chain of smaller theaters throughout the country. The government charged them with impeding and restraining competition through such practices as blocked booking, discrimination in favor of their own theaters, charging minimum admission prices, and protracting the intervals between successive showings of films, thereby affecting adversely and unfairly the independent producers and distributors. When the case finally reached the

Supreme Court in 1948, the decision of the Court was that production and exhibition be separated from each other. In 1952, after reorganization, the five companies became ten, consisting of five producers and five operating chains of theaters. Here, as stated previously with respect to monopoly, the Court felt that there was sufficient power and its abuse to bar effective competition, and hence the firms were held in violation of the law.

Mergers. The final effects of the Antimerger Act of 1950, which forbade the acquisition of assets as well as shares where the effect may be "substantially to lessen competition or tend to create a monopoly," remains to be seen. According to recent trends, the antitrust agencies will use their own judgment as to what constitutes a substantial lessening of competition or a tendency to create a monopoly in each particular situation, rather than wage an all-out war against mergers in general. Thus, in a recent case involving Pillsbury Mills, the company had acquired two other milling firms thereby raising its market share for flour-base mixes from 16 per cent to 45 per cent. In a preliminary hearing on the matter, the FTC concluded that there was *prima facie* evidence that competition might be substantially impaired (Pillsbury Mills, Inc., Docket No. 6000, Remand, Dec. 18, 1953). Similarly, recent requests by Bethlehem Steel to merge with Youngstown Sheet and Tube, which would increase its capacity to one fifth of the industry's, have been disapproved by the Department of Justice. On the other hand, in three instances involving a merger of automobile manufacturers—Kaiser and Willys, Nash and Hudson, and Packard and Studebaker—the enforcement agencies entered no complaint, probably believing that these mergers would increase competition with General Motors, Ford, and Chrysler. On balance, it seems that the antitrust agencies are concerned with distinguishing between mergers that will tend to lessen competition as against those aimed at product diversification, vertical integration, and the strengthening of weaker competitors.

Patents

The Constitution of the United States (Art. 1, Sec. 8, Par. 8) empowers Congress "To promote the progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries . . ." Though this power was not denied to the states, it came in time to be exercised solely by the federal government, and upon this authority the American patent and copyright system is based. In the present discussion our attention will be devoted exclusively to patents and their particular legal-economic aspects that are of concern to management.

A patent is an exclusive right conferred by a government on an inventor, for a limited time period. It authorizes the inventor to make, use, transfer, or withhold his invention, which he might do even without a patent, but it also gives him the right to exclude others or to admit them on his own terms, which he can only do with a patent. Patents are thus a method of promoting invention by granting temporary monopolies to inventors. But the patent system, it is held, has also been employed as a means of controlling output, dividing markets, and fixing prices of entire industries. Since these are perversions of the patent law which have a direct effect on competition, they have been subject to criticism by the antitrusters, and the courts have come increasingly in recent years to limit the scope and abuses of patent monopoly. Among the chief issues have been the standard of patentability, the right of nonuse by the patentee, the use of tying contracts, the employment of restrictive licenses, and the practices of cross-licensing and patent pooling. The recent trends based on court decisions in each of these areas may be outlined as follows.

Standard of Patentability. The chief standard employed by the courts is the so-called "flash of genius" test. Thus, in the Cuno Engineering Corporation case in 1941, involving the patentability of a wireless lighter, Justice Douglas, speaking for the Court, said that usefulness and novelty alone do "not necessarily make the device patentable. . . . The device must not only be 'new and useful,' it must also be an 'invention' or 'discovery.'. . . The new device, however useful it may be, must reveal the *flash of creative genius*, not merely the skill of the calling. If it fails, it has not established its right to a private grant on the public domain" (314 U.S. 84, 91; italics supplied).

The "flash of genius" test has been criticized as resting on the subjective judgment of the Court, and as not taking sufficient recognition of inventions that are the product of teams rather than individuals, especially in large corporations. In response to these arguments, Congress passed the Patent Act of 1952 which provides that a formula, method, or device, in order to be patentable, must be "new" in that it must be unknown to the public prior to the patent application, or it must be "useful" in that it evidences a substantial degree of technical advance in the object invented or in the process of producing something. But the courts have not found in the Act an adequate definition of "invention" and continue to rely on case law and their own judgment in determining what constitutes an invention (in re: *United Mattress Machinery Co.* vs. *Handy Button Machine Co.*, 98 USPQ 296, 299 [1953]). It appears, therefore, that the "flash of genius" test, tempered perhaps by the political and economic attitudes of the courts with respect to the public interest, will be the chief criterion of patentability at least in the foreseeable future.

Right of Nonuse. The right of a patentee to withhold an invention from use has been upheld by the courts. In numerous cases tried during the past sixty or so years, the courts have viewed a patent as a form of private property and hence have upheld the patentee's right to refuse putting it to use. In response, it has been argued by some that a patent is a privilege and not a right, that the practice of nonuse may result in retarding technological progress and economic development, and hence that the courts should exercise more judgment and discretion in such cases. And even the courts in recent years have spoken of patents as privileges contingent upon the enhancement of public welfare (in re: *Special Equipment Co* vs. *Coe*, 324 U.S. 730 [1945]). But the right of nonuse appears nevertheless to be supported by the law, for as stated by the Supreme Court in the Hartford Empire case in 1945: "A patent owner is not . . . under any obligation to see that the public acquires the free right to use the invention. He has no obligation either to use it or to grant its use to others" (323 U.S. 386).

Tying Contracts. These are viewed by the antitrusters as attempts to extend the scope of monopoly beyond the limits of a patent grant, and the courts have upheld this view by striking down consistently and repeatedly all such agreements. Two common forms of tying contracts that have been held illegal may be noted:

1. *Attempts by the patentee to prevent a competitor from selling an unpatented product in a patented combination.* In a case tried in 1944 involving Minneapolis-Honeywell against the Mercoid Corporation because the latter had sold an unpatented switch for use in connection with a patented combination of thermostats for controlling furnace heat, the Court found no patent infringement. It held that Honeywell's attempt to extend the scope of its patent was illegal; and in the words of Justice Douglas speaking for the Court, "An unpatented part of a combination patent is no more entitled to monopolistic protection than any other unpatented device" (330 U.S. 680).

2. *Attempts by the patent holder to require in the license contract that the licensee purchase other products from the patentee.* Before the passage of the Clayton Act such tying contracts were upheld by the courts, some well-known examples being the A. B. Dick case in 1912 and the United Shoe Machinery case in 1913. But since the passing of the Clayton Act, which outlaws such contracts in Section 3, a number of tying contracts have been struck down by the courts where such agreements were found substantially to lessen competition within the meaning of the Act (in re: *Lord* vs. *Radio Corp. of America*, 278 U.S. 648; *International Business Machines Corp.* vs. *U.S.*, 298 U.S. 131; and *International Salt Co.* vs. *U.S.*, 332 U.S. 392). It appears, therefore, that the trend of the courts is to disallow tying clauses of any kind, regardless of circumstances, where the effect is to extend the scope of a patent monopoly.[1]

Restrictive Licensing. This refers to the practice of licensing patents among competitors with certain restrictions imposed. Typically, the restrictions may include the patentee fixing the geographic area of the licensee, his level of output, or the price he may charge in selling the patented goods. Usually such licensing is motivated by considerations of reciprocal favor (e.g., the exchange of patents among competitors) or perhaps performed for the purpose of minimizing the incentive of the licensee to develop an alternative process. In any case, since it is the legality of particular practices that concerns us here, the following trends may be noted.

---

[1] Thus an extreme example existed with respect to Eastman Kodak prior to 1954. The company sold amateur color film at a price which included the charge for finishing, thereby tying the sale of the film itself to the business of providing finishing services. In 1954 the company signed a decree, agreeing to sell the film alone and thus admit competitors to the finishing business.

1. The right of a patentee to fix the licensee's prices on patented products has been and still is upheld by the courts. The leading case in this respect, decided in 1926, concerned the question of whether General Electric could, under its basic patents on the electric lamp, fix the prices charged by Westinghouse, the licensee. The Supreme Court answered in the affirmative (272 U.S. 490). It should be observed, however, that this case involved a *single* patentee, a *single* licensee, and a *single* product.

2. The right of the patentee to fix the prices charged for unpatented products made by patented processes (e.g., a patented machine) is very doubtful (in re: *Barber Coleman Co.* vs. *National Tool Co.*, 136 F 2d 339 [1943]; and *Cummer Graham Co.* vs. *Straight Side Basket Corp.*, 142 F 2d 646 [1944]).

3. In contrast with the General Electric case cited above, the use of restrictive licensing is illegal when employed for the purpose of eliminating competition among many licensees. In the Gypsum case decided in 1948, involving a price-fixing arrangement among the licensed producers of gypsum wallboard, the Supreme Court upheld the government's charge. In the words of Justice Reed: "Lawful acts may become unlawful when taken in concert," and therefore "the General Electric case affords no cloak," no precedent, "in this case" (*U.S.* vs. *U.S. Gypsum Co.*, 333 U.S. 364,400). Thus there is now a sharp restriction as to the extent to which a patent owner may license his patent. On the basis of this and the Line Material case (333 U.S. 287; 1948), when each of several licensees accepts restrictive terms on condition or with the knowledge that others will do likewise, they are committing a conspiracy in restraint of trade in the opinion of the Court and hence are guilty of violating the law.

Cross-Licensing and Patent Pooling. These are not held to be illegal as such, but they generally are declared illegal when, in the eyes of the courts, they are used as a means of eliminating competition among patent owners and licensees. But what constitutes elimination of competition? In the Hartford Empire case, decided in 1945 (323 U.S. 386), it was held that Hartford employed the patents in its pool to dominate completely the glass container industry, curtail output, divide markets, and fix prices through restrictive licenses, and therefore this was unlawful conspiracy. In the National Lead case in 1947 (332 U.S. 319), a cross-licensing agreement that divided markets and fixed the prices of titanium pigment was also declared illegal. And in the Line Material case of 1948, cited previously, the Court was most emphatic in its denouncement of a cross-licensing arrangement that fixed the price of fuse cutouts used in electric circuits. In general, it appears that although patent pooling *per se* is not illegal (the automobile industry being frequently cited as an outstanding example of successful and desirable patent pooling), the courts will declare it illegal when it seems to be abused. And the courts will tend to declare that abuse exists when either the pool is restricted to certain competitors or available only at excessive royalty payments, or when the pool is used as a device to cross-license competitors for the purpose of fixing prices and allocating markets.

Concentration of Patent Ownership. Patent concentration in the hands of a single firm has also come under consideration in recent years. Prior to another United Shoe Machinery case in 1953 (110 F. Supp. 295), the ownership of many patents by a single company was held to be legal (in re: *Transparent Wrap Machine Corp.* vs. *Stokes & Smith Co.*, 329 U.S. 637 [1947]; *Automatic Radio Mfg. Co.*, vs. *Hazeltine Research, Inc.*, 339 U.S. 827 [1950]). But the United case appears to represent what may be a definite turning point in the trend of the Court. Thus, the Court found that the Company: (1) had almost 4,000 patents, about 95 per cent of which came from its own research and the remainder purchased from others; (2) put about a third of these patents to use; (3) had not abused its patents by suppressing them or by using them as a threat over competitors; (4) had not been offered or asked to grant licenses, but had not refused to do so; (5) had not resorted to litigation as a means of harassing competitors but instead acted in good faith in bringing infringement suits; (6) had leased rather than sold its machines, and in a manner so as to discriminate against customers who might install competing machines; (7) required lessees to use the machines at full capacity in the manufacture of shoes; and (8) required that lessees purchase United's supplies and services along with the leasing of machines. None of these policies, the court held, was illegal *per se*, but their combined effect, in view of United's dominant position in the industry, was sufficient to constitute monopolization and hence the firm was held in violation of Section 2 of the Sherman Act.

Whether the decision of the Court in the Shoe Machinery case is really the beginning of a new trend is not yet certain. At the present time several cases are pending concerning the concentration of patents through research, assignment, and purchase, and the outcome of these cases, which involve some major corporations, will provide a stronger base upon which to predict future court decisions. This much is certain, however: in recent years, some strong remedies have been used by the courts against patent holders who have been declared in violation of the antitrust laws. These remedies, which are now quite typical include: compulsory licensing, sometimes on a royalty-free basis for a company's existing patents, and on a reasonable royalty basis for future patents; and the provision of necessary know-how, in the form of detailed written manuals and even technical consultants, available at nominal charges, to licensees and competitors. Thus Eastman Kodak agreed to provide other color-film finishers with up-to-date manuals on its processing technology and to provide technical representatives to assist competitors in using the methods described. In a number of other cases involving Standard Oil of New Jersey, the Aluminum Company of America, Merck & Co., A. B. Dick, Libbey-Owens-Ford, Owens-Corning Fiberglas, American Can, and General Electric, as well as about twenty-five other firms, somewhat similar provisions have been arrived at since the 'forties. Hundreds of patents involving a wide variety of manufacturing areas have thus been freed, and it is to be expected that the courts will continue to move in this direction in future years.

Trade Marks

The purpose of a trade mark, as originally conceived, was to identify the origin or ownership of a product. In an economic sense, however, managements have come to look upon trade marks as a strategical device for establishing product differentiation and, through advertising, strong consumer preference. In this way firms have sometimes been able to establish a degree of market entrenchment that has remained substantially unrivaled for as long as several decades. Moreover, by establishing product differentiation through trade marks, firms have exploited this advantage in various ways with the aim of enhancing long-run profits. Five examples may be noted in view of their antitrust significance.

1. Price discrimination has been implemented by the use of trade marks. As mentioned in the previous chapter, Rohm & Haas sold methyl methacrylate as Lucite and Crystalite to manufacturers at 85 cents per pound, and as Vernonite and Crystalex to dentists at $45 per pound. The decision, rendered in 1948, was against the company for using trade marks in this discriminatory manner. (*U.S.* vs. *Rohm & Haas,* Civil Action No. 9068, Dist. Ct. of the U.S., Eastern Dist. of Pa.)

2. Output control has been accomplished through the use of trade marks. U.S. Pipe and Foundry licensed companies to produce under its patents at graduated royalty rates on condition that they stamp their products with the trade name "de Lavaud." The decision, rendered in 1948, was against the company for using a trade mark in controlling output. (*U.S.* vs. *U.S. Pipe and Foundry Co.,* Civil Action No. 10772, Dist. Ct. of the U.S., Dist. of N.J.)

3. Exclusive markets have been attained through the use of trade marks. General Electric was able to persuade procurement agencies to establish specifications requiring the use of Mazda bulbs. It licensed Westinghouse to use the name but denied its other licensees the same right. The decision against General Electric was rendered in 1949, on the

grounds that the Company had used the trade mark as a device for excluding competitors from markets. (*U.S.* vs. *General Eelectric Co.*, 82 F. Supp. 753.)

4. Market sharing by international cartels has been accomplished through the use of trade marks. The procedure is somewhat as follows. A trade mark is advertised throughout the world and each cartel member is granted the exclusive right to use it in his own territory. If a member oversteps his market boundary, he is driven back by an infringement suit. Trade names that provide examples of such regional monopolies include Mazda, Mimeograph, Merck, and Timken, and the trade marks of General Storage Battery, New Jersey Zinc, American Bosch, and S.K.F. Industries. In a number of cases tried during the late 'forties, the courts found such arrangements to be in violation of the Sherman Act. In the Timken Roller Bearings case of 1949 (83 F. Supp. 294), it rejected the licensing of trade marks as a defense; in the Merck & Co. case of 1945 (Civil Action No. 3159, Dist. of N.J.), it canceled trade marks and enjoined their renewal; and in the Electric Storage Battery case in 1947 (Civil Action No. 31–225, Southern Dist. of N.Y.), it forbade cartel members the right to grant their foreign partners exclusive trade mark rights abroad, to sell in American markets, and to interfere with American imports. In short, where trade marks have been employed to implement market sharing arrangements by cartels, the courts have usually upheld the government with stringent remedies, and probably will continue to do so. Where trade marks have supported purely domestic monopolies, however, the government has tread more lightly. Thus in the American Tobacco case of 1946 cited earlier, the leading manufacturers of cigarettes were found guilty of violating the Sherman Act. Nevertheless, the government, unlike in other cases, did not request dissolution, probably because it would have resulted in the destruction of property values, i.e., brand names (Camel, Chesterfield, Lucky Strike) that are now worth millions of dollars. It appears, therefore, that the exclusive right to a name that has been widely advertised may continue to be held as an important consideration in the future applications of the antitrust laws.

5. Finally, resale price maintenance has been implemented by the use of trade marks, even where patents and copyrights have failed. Although contracts which maintain the resale price of trade-marked goods were held to be unlawful as early as 1911 in the Dr. Miles case (220 U.S. 373), they have subsequently been legalized and will be discussed more fully at a later point in this section.

Tying Contracts and Exclusive Dealing

Tying contracts and exclusive dealings have sometimes been used, as shown earlier, to obtain and extend a position of monopoly. In the opinion of the antitrusters, such agreements affect the ability of producers to compete with one another in obtaining access to markets, and the ability of distributors to compete with one another in the purchase and resale of goods. And in most cases, the courts have upheld the government in its view by striking down such arrangements, its decision usually hinging on whether it believed the effect was "to substantially lessen competition or tend to create a monopoly." But when is competition substantially lessened? Congress never explained this when it passed the Clayton Act, and it has been left to the courts to decide in each case. Opinions have varied and, in general, it cannot be predicted that exclusive arrangements will be outlawed *per se*, despite the pressure exerted by the FTC. At best, all that can be said is that the Commission, on the basis of recent behavior, is confining its orders to cases in which it believes it can actually show substantial injury or the probability of such injury to competition. Perhaps this is due to the courts which, in recent years, have dismissed a suit or two because of insufficient evidence, declaring that the use of coercive methods to secure exclusive dealerships is unlawful, but exclusive dealerships as such are not illegal (*U.S.* vs. *J. I. Case Co.*, 101 F. Supp. 856 [1951]). This, in essence, is where the issue now stands.

Price Discrimination—General Legality

With respect to price discrimination, the antitrust laws have been applied primarily to two classes of practices: the first is discount structures; the second is delivered pricing systems. Both of these are considered in the following subsections after a discussion of the general legality of price discrimination.

The Robinson-Patman Act, which amended Section 2 of the Clayton Act dealing with price discrimination, made the following practices illegal: (1) charging different prices to different buyers on sales that are otherwise identical; (2) selling at different prices in different parts of the country "for the purpose of destroying competition or eliminating a competitor"; (3) selling "at unreasonably low prices" where the purpose is to destroy competition or a competitor; (4) discriminating in price; (5) paying brokerage commissions to buyers or to intermediaries under their control; and (6) granting allowances, services, or facilities by sellers to buyers, whether for services rendered by the buyer or not, that are "not accorded to all purchasers on proportionally equal terms."

The first of these offenses is too narrowly defined to be of much practical importance; the second is definite; the third is too vague and difficult to enforce: i.e., when are prices "unreasonably low"? These first three provisions constitute the criminal portion of the Act and have been relatively insignificant. It is in its civil aspects or in the last three provisions that the Act has been of major importance, and these will concern us here.

Under the Act, price discrimination is illegal not only where the effect is "to substantially lessen competition or tend to create a monopoly," as in the Clayton Act, but also where it may be "to injure, destroy, or prevent competition with any person who either grants or knowingly receives the benefit of such discrimination, or with the customers of either of them." The Act thus makes *injury to competitors* the test of illegality. It also, however, allows the seller charged with discrimination to offer two defenses: (1) that the differentials in price "make only due allowance for the differences in cost of manufacture, sale, or delivery . . ."; and (2) that the lower price "was made in good faith to meet," not competition, as in the Clayton Act, but "an equally low price of a competitor," and not to undercut it. That is, discriminatory price cutting to "meet" competition is legal; discriminatory price cutting to "beat" competition is illegal. Finally, the Act also permits the FTC to set limits on quantity discounts, even though justified by cost differences, "where it finds that available purchasers in greater quantities are so few as to render differentials on account thereof unjustly discriminatory or promotive of monopoly," and it makes it illegal for buyers "knowingly to induce or receive" a prohibited discrimination in price.

Evaluation. What economic significance may be attached to the Robinson-Patman Act? At least two classes of problems, involving measurement and policy, may be considered.

First, the underlying principle of the Act is that differences in price should be proportionate to differences in cost. However, anyone familiar with cost accounting knows that although differences in costs resulting from alternative methods of selling and delivery may not be too difficult to measure, the measurement of costs resulting from manufacturing creates problems of accounting theory that are open to various interpretations. The allocation of overhead provides a typical example. Should lower unit overhead costs be attributed to larger orders? Should more overhead be charged against goods in periods of full capacity, and less when there is idle capacity? Should only the extra costs of filling an order be considered in pricing, on the assumption that overhead is recovered on other sales, or should the overhead be allocated uniformly and hence the total unit costs be estimated by dividing total costs by total output (as in economic theory)? The Act provides no answers, and cost accountants themselves are not in complete agreement on these and other issues.

Second, policy-wise, the law is inherently contradictory in that it sometimes outlaws discrimination, sometimes permits it, and sometimes even requires it. Thus, sellers may legally discriminate among consumers, and among noncompeting business buyers in the channel of distribution (e.g., manufacturers, wholesalers, and retailers). They may also charge identical prices where costs differ, which is also discriminatory, or give discounts that do not reflect real cost differences. And they may discriminate *against* the firm that buys in quantity. However, they must not discriminate *in favor* of the firm that buys in quantity, and they must discriminate when they deny a broker's commission to a buyer performing a broker's function, or when allowances or services are denied though they would pay for themselves by promoting sales, or when the Commission sets limits on quantity discounts even though the discounts reflect real differences in cost.

In view of these considerations, what are the consequences of the Act? Though the results are not measurable, this much seems plausible. First, the requirement that discounts be justified by actual rather than potential cost differences has probably discouraged what would otherwise have been profitable price reductions, and hence has possibly impeded gains in efficiency and orderly business growth. Second, by making it illegal to set lower prices in particular markets in order to test for the possibility of increasing sales, the Act has probably prevented many price reductions that would eventually have been generalized. And third, by withdrawing from the mass distributor certain advantages that he formerly obtained, his costs and prices are probably higher than they otherwise might have been. However, the law has also had at least two positive effects, in that concessions previously given in one form may now be given in another. Thus, discounts as large or larger than those once granted may still be justified by cost differences; and a mass distributor may buy a plant's entire output or else manufacture the product himself. Costs may thus be reduced without involving discrimination.

What all this adds up to is this. The authors of the Clayton Act were concerned with the effects of discrimination on competing sellers; the authors of the Robinson-Patman Act were concerned with the effects of discrimination on competing buyers. The latter law was designed to reduce the buying advantages of chains and mass distributors; it aimed, not at eliminating discrimination in general, but at preventing discrimination in favor of larger buyers and permitting or requiring it in favor of smaller ones. And it has been applied in a manner that has served not only to handicap the chains, but also to check the advantages obtained by agencies that buy collectively for independent firms. In short, compared with the Clayton Act, it was concerned more with the maintenance of small competitors and less with the maintenance of competition. But on balance, whether the effect has been actually to strengthen or weaken competition, it is not possible to say.[2]

With this as a background, we now turn our attention to the legal status of discounts and delivered pricing as the most important specific areas of price discrimination, and the recent court trends in each.

Price Discrimination—Discounts

The courts have, under the Robinson-Patman Act as under the Clayton Act, upheld the government in cases involving local price discrimination. Discounts, however, have been treated less systematically, depending primarily on the form of the discount. In the following paragraphs the current legal status of various kinds of discounts are surveyed, along with enforcement problems, in the light of recent court trends.

---

[2] Professor Wilcox believes that, if anything, it helped the A & P by forcing it into the supermarket business. (C. Wilcox, *Public Policies Toward Business*, pp. 187–88.)

Brokerage and Allowances. The illegality of brokerage payments has been consistently upheld in a number of different circumstances. In fact, about half the orders issued by the Commission under the Robinson-Patman Act have been to prohibit such payments. Two typical classes of circumstances may be noted:

1. Where brokerage has been given as an advantage to a single buyer, it has been held illegal. Thus, in the A & P case of 1940, the company was charged with receiving commissions which were granted to it in the form of quantity discounts and price reductions. The company's defense was to justify the procedure in terms of cost, claiming that its agents in the field served not only in a purchasing capacity for A & P, but that they also saved sellers a brokerage cost by advising the latter group on how to dispose of their surpluses. But the court rejected this defense on the grounds that such payments were, under the Robinson-Patman Act, unqualifiedly illegal (308 U.S. 625). And in the Webb-Crawford case in 1940 (210 U.S. 638), involving the owners of a wholesale grocery firm who were also partners in a brokerage organization, the court held that the collection of commissions through a dummy firm was illegal.

2. Where brokerage has been passed on to the benefit of smaller firms, it has also been held illegal. Examples include the Oliver Bros. case of 1939 (102 F. 2d 763), a firm which sold marketing information to its clients, and passed its commissions on to them in the form of lower prices; and the Modern Marketing Service case of 1945, a purchasing firm for wholesale and retail grocers which passed on to them the commissions it received from suppliers (149 E 2d 970). In these and similar cases, some involving cooperatives, payments that were deemed helpful to small independents, and not to mass distributors, were nevertheless prohibited.

Allowances and services have been given much less attention, possibly because of the ambiguity of the law. The Robinson-Patman Act states that allowances and services must be made available to all buyers "on proportionally equal terms," but it does not state a criterion for proportionality nor has the Commission defined one. Several possibilities include: proportionality to dollar volume of sales; proportionality to the buyer's cost of the services rendered to him; and proportionality to the value of such services to the seller. Although no one of these has been consistently applied, the first, if any, would appear to be the most likely one. In any event, only a few cases have ever reached the courts, but on the basis of these and on several complaints issued by the Commission, the interpretation of "allowances" seems to be this: they must not be secretly rendered, but must be publicly announced; their terms must not be such as to confine them to a few large buyers, but must be available to all; they must be made only for services actually rendered; and they must not be excessively greater than their cost to distributors or their value to manufacturers (in re: *Corn Products Refining Co.* vs. *FTC*, 324 U.S. 726 [1945]; *Elizabeth Arden, Inc.* vs. *FTC*, 331 U.S. 806 [1947]). In general, therefore, it appears that the courts, as in the past, will continue to follow the FTC in the latter's strict interpretation of the sections of the Act dealing with brokerage and allowances.

Enforcement Problems. The prohibitions contained in the Robinson-Patman Act concerning brokerage and allowances make such discounts unqualifiedly illegal. But other forms of discounts are illegal only where the effect "*may be substantially* to lessen . . . or to injure, destroy, or prevent competition . . ." (italics supplied), or where they do not "make only due allowance for differences in the cost of manufacture, sale or delivery resulting from the differing methods or quantities in which . . . commodities . . . are sold or delivered." The test of illegality thus hinges on the word "substantially" in the first quotation, although the word "may" is also of considerable significance. Only in relatively recent years have the courts interpreted these words to mean that a *reasonable probability* of injury to competition, rather than mere possibility, must be affirmatively proven. Thus in the Minneapolis-Honeywell case in 1951, the Court stated, "We construe the Act to require substantial, not trivial or sporadic, interference with competition to establish the violation of its mandate" (191 F 2d 786), and this position was upheld by the Supreme Court in 1952 (344 U.S. 206). For some time prior to this, as exemplified by the Morton Salt case in 1948, reasonable "possibility," rather than "probability," of injury to competition was sufficient to establish guilt (334 U.S. 37). This change in court interpretation is important with respect to enforcement of the law: "probability" at least requires some proof, while "possibility" could conceivably be assumed. And the Commission has also adopted this standard, holding that a "reasonable probability" of injury must be proven.

As for "due allowance for differences in cost," the Commission has provided no guides of accepted cost accounting principles and, until recently, has frequently rejected cost estimates offered in defense of price differences. In general, its policy has been to: (1) permit discounts justified by savings in selling and delivery costs; (2) reject discounts based on savings in manufacturing costs; (3) establish average total cost of an order (total expenditures ÷ total output) rather than marginal cost (added expenditure ÷ added output) as its standard or, in other words, require a uniform allocation of overhead to all units sold, thereby denying the role of incremental cost as a management guide in production and pricing; and (4) put the burden of proof on the seller as to whether there is really a cost difference. In short, it has instituted its own brand of economics, much of which is counter to established and accepted principles, by sometimes rejecting demonstrated savings as being inapplicable to the particular prices involved, sometimes accepting large differences in costs as justifying small differences in price, and sometimes rejecting whole accounts on the grounds that the savings claimed appeared to be excessive and speculative.[3] In 1953, however, under a new chairman, the Commission began what may be a more liberal policy, by attempting to establish sounder accounting and economic guideposts with respect to "due allowance." On the basis of official statements, it appears that the Commission will give more consideration to the intricate and complex problems of cost analysis than it has in the past, and will exercise a more reasonable attitude in future cases.

Quantity Discounts. These have never been held, either by the Commission or by any court, to be illegal *per se*. In fact, in the Bruce's Juices case in 1947, Justice Jackson, speaking for the majority of the court, said: "The economic effects on competition of such discounts are for the Federal Trade Commission to judge. . . . It would be a far-reaching decision to outlaw all quantity discounts. Courts should not rush in where Congress fears to tread" (330 U.S. 743, 746). And the Commission, possessing the power to judge the effects of discounts on competition, has frequently exercised that power. It has prohibited discounts where it could not find sufficient savings in delivery, selling, or production costs to justify a price difference, and it has forbidden cumulative discounts, i.e., discounts on purchases over periods of time, claiming that such purchases do not evidence a reduction in seller's costs (in re: Standard Brands, 30 FTC 1117 [1940]). Further, the Commission has sometimes ruled that discounts at a given time and place may be justified by differences in cost, as with the Kraft-Pheenix Cheese Corp. in 1937 (25 FTC 537) and the American Optical Co. in 1939 (28 FTC 169), and sometimes not, as in the landmark Morton Salt Co. case in 1948 (334 U.S. 47). Morton's prices, it was charged, were unjustly discriminatory and injurious to competition because its discount structure, though in principle available to all the company's purchasers, was in practice available only to the largest few. That is, the company's prices ranged from $1.60 to $1.35 per case for quantities ranging from less-than-carload lots to 50,000 cases, purchased within a year. The evidence showed that only a few large grocery chains could buy enough within a year to take advantage of the $1.35 price, while independent retailers had to buy from wholesalers who paid around $1.50. Both the Commission and the Court held this as substantially injurious to competition, and the company has since withdrawn all discounts on quantity buying.

The practical outcome of the Morton Salt decision should be noted, in view of its unfortunate economic consequences as well as the problems it has left unresolved. Manifestly, it is within the discretion of the Commission to decide whether a price difference, no matter how small, is discriminatory or not. The burden of proof—and a heavy burden it frequently is—then passes to the seller, who must justify the difference either by "due allowance" for differences in cost, or by "good faith" to meet, but not undercut, a competitor. By thus failing to distinguish between a price difference and a price discrimination, the Commission can conceivably declare any quantity discount, no matter how small, a violation of the law.

---

[3] *Cf.* Wilcox, *ibid.*, p. 193, and M. A. Adelman, "The Consistency of the Robinson-Patman Act," *Stanford Law Review*, Vol. 6 (1953), pp. 3–22.

Of what significance is this? The answer is that the power held by the Commission may result in the elimination of quantity discounts by many sellers, even though such discounts, after long and expensive litigation in the courts, might have been upheld as nondiscriminatory. The outcome is twofold: quantity buyers who otherwise might have passed their savings on to consumers are unable to do so, at least in the short run; and the public, through higher prices, preserves small business on the questionable assumption that such preservation is in and of itself desirable.

Functional Discounts. In general, the attitude of the Commission with respect to trade (or functional) discounts is that they are legal, and that discrimination between buyers is not unlawful, provided that the buyers are not in competition with one another. Is it possible, therefore, for sellers to evade the law relating to quantity discounts by establishing special customer classes for the purpose of granting discounts that cannot be justified by cost differences, or, in other words, by cloaking unjustifiable quantity discounts as functional discounts? The answer is no, not readily. As mentioned in the previous chapter, the Commission has stated the conditions of classification, which are as follows:

1. Buyers must be classified according to their strict nature or level of operations. Different types of customers who nevertheless are at the same level, e.g., chain stores, independent retailers, and mail-order houses, must be placed in the same class, and the discounts granted must not exceed the cost savings (in re: American Oil Co., 29 FTC 857 [1939]; Sherwin-Williams, 36 FTC 25 [1943]).

2. For split-function customers, such as a dealer who is both a wholesaler and a retailer and thus performs two functions, the discount allowed for any function must be applied only to the portion of the order for which that function alone is performed (in re: Standard Oil Co. of Indiana, 41 FTC 263 [1945], and 43 FTC 56 [1946]). This rule, however, has two shortcomings: (*a*) it is difficult to enforce because the seller can only take the buyer's word as to which quantities will be employed in each function, and the buyer may naturally tend to overstate the quantity on which the larger discount will apply; (*b*) it denies the split-function dealer the discount he rightfully deserves for performing the wholesale function on that part of the goods that he retails himself. The rule thus attempts to maintain a rigid stratification of functions in distribution, by trying to prevent dealers who perform both wholesale and retail functions from reducing the retail price of their goods. And the Commission, it appears, is on fairly solid ground, despite the dubious assumption on which it operates

that everything that "injures" (hurts) competitors "injures" (hinders) competition.[4]

Conclusion. Two defenses are available to the seller who is charged with illegal discrimination: one of these is to show that the differences in his price made due allowance for differences in cost; the other is that the lower price was made in good faith to meet, but not undercut, the price of a competitor. From what has been said in the previous paragraphs, it is clear that the Federal Trade Commission has seriously impaired both of these defenses. Evidently, the FTC's policy is to lend encouragement to "soft" competition and to frown on "hard" competition as required by the Sherman Act. Under a policy of hard competition, price discrimination would still be controlled, but there would be some important differences. Since the application of the law hinges on two considerations, namely (1) the test of illegality, and (2) the respondent's defense, the following amendments to the Robinson-Patman Act would seem advisable as an initial step in the right direction.

1. The test of illegality should be changed from injury to a competitor to injury to competition in general, with emphasis on the probability rather than the mere possibility of injury being shown. As implied previously, not everything that hurts competition hinders competition. A price discrimination may frequently hurt a competitor, as does any price cut, but whether it hinders competition in general is much less certain. There is as much if not a greater likelihood that it promotes competition rather than hinders it.
2. The two defenses, i.e., the "cost defense" and the "good faith defense," should be reconsidered.

Cost Defense. The rule that differences in price should make due allowance for differences in cost, should be retained. However, accounting rules should be established which recognize savings in manufacturing as well as in selling and delivery, and which utilize incremental rather than average costs as a criterion. The employment of incremental cost as a guide would be in accord with management's use of this concept as a choice indicator in decision problems, and hence would reflect more accurately the significance of a particular act. Further, the provisions of the Act relating to brokerage, allowances, and services should be revised so as to permit such concessions, provided they are related to the cost and the value of the services involved so that they are not used to evade the rules governing quantity discounts. And finally, the Commission's power to limit quantity discounts, even when justified by lower costs, should be repealed.

---

[4] The leading case in this respect is *Standard Oil Co.* (*Ind.*) vs. *FTC,* 340 U.S. 231 (1951), and the Commission's Modified Order, Docket No. 4389 (1953), concerning Ned's in Detroit, a jobber and cut-rate retailer of gasoline. The Court reversed the Commission's ruling, although on other grounds. For a concise analysis, see Wilcox, p. 196–98, and G. Stocking and Watkins, *Monopoly and Free Enterprise,* p. 374, note 92.

Good Faith Defense. The good faith defense should be retained as a means of justifying sporadic price cuts made in order to meet the lower prices of competitors, and not as a protection for systematic discrimination such as occurs in delivered pricing systems (discussed next).

With these amendments, the Robinson-Patman Act would improve the rivalry among business firms, and the maintenance of competition would thus become more nearly self-enforcing. As it stands now, the Act is concerned more with the survival of small competitors than with maintaining competition. Thus intended, it does not prevent discrimination in general, but merely prevents discrimination in favor of larger buyers and permits or even requires it in favor of smaller ones.

Price Discrimination—Delivered Pricing

The preceding discussion of discrimination was rooted primarily in the economic functions performed by purchasers. In the present subsection we turn our attention to another form of discrimination—geographic price discrimination—which arises because of the particular locational differences that exist between buyers and sellers. Such differences may result in delivered pricing systems, which means essentially that the price to the buyer includes not only the cost of the goods themselves but a delivery charge as well. The result of this is that the seller's mill net will vary depending on the amount of freight charges he absorbs himself, and by thus accepting varying net returns on sales to different customers, he is discriminating between customers.

How important is this discrimination with respect to antitrust? The answer is that it will be significant: (1) where the transportation cost is a large proportion of the final price, or in other words where the value-transport ratio is low; (2) where sellers are pricing on the basis of mutual understanding, either tacit or explicit, so that the effect is to restrain competition. Geographic discriminatory pricing may exist under the opposite of these conditions, i.e., a high value-transport ratio and independence of sellers actions; it may also occur with either one of the conditions; but it is mainly when both exist, and especially the second, that the antitrusters have been particularly concerned.

The economic nature of delivered pricing systems, such as basing points, zones, and freight equalization, have been outlined in the previous chapter. In the following paragraphs we shall be concerned primarily with the legality of delivered pricing systems as evidenced by recent cases and court decisions. Since most of the litigation has centered on basing point systems and their ramifications, the discussion below will be oriented around these practices. The procedure followed will be to outline first the implications of the basing point controversy; second, the alternatives to basing point pricing; and third, the legal status of basing point systems.

The Basing Point Controversy. The basing point system has been debated in economic literature and court cases since the time of World War I. Out of the complicated mass of facts and interpretation, two distinct schools of thought have emerged. One of these is the opponent group, composed of the Federal Trade Commission and a large number of academic economists, who hold that basing point systems are monopolistic and the result of collusion, and should be generally outlawed; the other is the proponent group, consisting of executives of basing point industries and some business and academic economists, who argue that such systems are competitive, that they emerge naturally in certain (oligopolistic) industries, and that to outlaw them would result in less desirable pricing systems. Let us examine these arguments and various related considerations somewhat more closely.

First, that the basing point system contains certain characteristics which, by economic criteria, would classify it as monopolistic, is not easily denied. There are mountains of evidence, consisting of data, statements by industry executives, and the like, to support the government's contention that identical and stable prices, as have occurred under basing point systems, could not have occurred by mere chance, nor by the "free play of market forces" as that expression is commonly understood in the theory of perfect competition. Benjamin Fairless, of U.S. Steel, himself admitted before the TNEC in 1940 that by full adherence to the basing point system, ". . . there wouldn't be any competition in the steel industry. It would be a one-price industry, pure and simple."[5] Nor is it very meaningful to say, as has been said, that the system has not always been adhered to, especially in periods of slack demand when price cutting and secret concessions are common. For the evidence indicates, at least for steel and cement, that the system was fully or almost fully adhered to in prosperity periods when operations were near or at full capacity and the pressure for price cutting almost nonexistent.

Second, that the basing point system "just growed like Topsy" because of the oligopolistic nature of the industries employing it is also difficult to establish. And moreover, whether it did nor did not, is irrelevant. On the one hand, there is sufficient documented evidence to indicate that the system, at least in the steel industry, was not natural or spontaneous. There were the famous Gary dinners between 1906 and 1911 in which Judge Elbert Gary, then president of U.S. Steel, presided over discussions with competitors as to pricing policy. And subsequently, the evidence for several basing point industries shows that organizations were formed, rules established, and meetings held, with violators of the system punished by fines, price raiding, and the like. If the system were genuinely spontaneous —an outgrowth of natural rather than artificial causes—as Clark, De Chazeau, Smithies, and some other economists have held, it seems doubtful that the industries would have gone to such lengths to preserve it. On the other hand, as stated above, whether the system emerged naturally or artificially is of no relevance, and the argument really amounts to much ado about nothing. The facts are that basing point systems are characteristic of oligopolistic industries, that such industries are dominant in the American economy, and, therefore, that the alternatives to basing point pricing systems must be considered if the antitrust laws and their interpretation by the courts are to be constructive rather than destructive.

---

[5] *Hearings before the TNEC*, part 27 (January 26, 1940), p. 14, 172.

Alternatives to Basing Point Systems. Three major possibilities may be considered as alternatives to basing point pricing. The first and most extreme one is compulsory f.o.b. mill pricing; the second, which is an opposite extreme, is to permit systematic freight equalization; the third, representing a compromise, is f.o.b. pricing with sporadic freight absorption. All three may be evaluated briefly and from a realistic standpoint, in the light of recent antitrust cases and court decisions to be discussed in the next subsection.

Compulsory F.O.B. Mill Pricing. Proposed by those who are the severest critics of basing point pricing, this method would require that each seller charge his buyers a uniform price at the mill, with the buyer arranging for delivery. The advantages claimed for the method by its advocates are that it would reduce costs and prices by: (1) confining each mill to a regional market, and thereby eliminating cross-hauling and competitive salesmanship, (2) enabling the buyer to use the cheapest means of transport available, (3) relocating mills in more economic areas according to demand requirements, and (4) making for more active competition in general.

These arguments as an alternative to basing point pricing are not entirely valid. First, under compulsory f.o.b. pricing, each mill would have an exclusive market where its price plus freight was lowest, with competition occurring only at the boundary where markets overlapped. But this status quo would be only temporary. With the first slackening of demand, intermarket penetration would occur via competitive price cutting, rather than (illegal) freight absorption, thus creating a pushball type of rivalry which, according to oligopoly theory, would undoubtedly end in market-sharing and price-fixing agreements.

Second, it is not certain that f.o.b. pricing would promote a more economical shift of productive capacity from surplus to shortage areas. If it did, it would be only at considerable cost in terms of the disemployment of resources (plant, equipment, labor, etc.) that would result. Further, such a shift would likely make for less, rather than more, competitors, because most new plants would probably be built by existing firms rather than new ones, and because smaller firms would find survival virtually impossible without freight absorption.

Finally, it is doubtful whether the prohibition of all freight absorption is advisable. This practice is economical when capacity is idle, and it serves as a means for increased rivalry since discrimination can and has been used not only to meet competitor's prices, but to undercut them as well. On all three counts, therefore, compulsory f.o.b. pricing would seem to be a wrong method of attack on basing point systems.

Systematic Freight Equalization. The second of the two extreme alternatives, systematic freight equalization, would amount to abolishing phantom freight by making every mill a basing point and requiring freight to be computed by the actual transportation method used, or by the cheapest of those available. Sellers would be permitted to equalize freight and prices systematically (consistently) at destinations, and thus discriminate among their customers. The adoption of this type of pricing policy would remove the traditional barriers to independent action as well as eliminate the artificial aids to price uniformity (e.g., common freight rate books) employed by sellers. But would this make for greater price differences? Not likely. The industry would still be characterized by oligopoly: some sellers would still await the action of a price leader before announcing their changes in base prices; each seller may still charge the lowest price quoted at any destination, thus resulting in identical prices; and the level of prices would still, therefore, be as high and as rigid as it was before. In short, although phantom freight will be eliminated, systematic discrimination through variations in the amount of freight absorbed would still be practiced, thus raising the question of whether a third alternative might not be preferble.

F.O.B. Prices with Sporadic Freight Absorption. This third alternative seems to be the most realistically feasible one in view of the legal aspects of basing points to be discussed next. Adoption of this compromise alternative would mean that buyers are given the option of taking delivery at the mill, and the seller can, at his own discretion, absorb freight in order to compete for a distant sale. The preference for this pricing policy is derived from the fact that it avoids the worst consequences of both basing point and compulsory f.o.b. pricing, while incurring some of the advantages of both.

1. It eliminates the artificial obstacles to economy in the selection of plant locations and transportation methods, by permitting new plants to be constructed in shortage areas and old plants to follow the market by absorbing freight. Stranded idle capacity and disemployed resources, as would occur under compulsory f.o.b. pricing, are thus avoided.
2. It wastes less transportation resources than a basing point system, but does not eliminate such waste entirely because some cross-hauling would still exist.
3. It encourages a diversification of costs in areas where markets overlap by providing for different methods of delivery, and thereby promotes some further competition in price.
4. It reduces the tendency toward market sharing that results from a fixed pattern of base price differentials because it permits sellers to invade one another's markets by absorbing freight.

The adoption of this pricing method would thus serve as an entering wedge for competition while at the same time minimizing the tendency toward monopoly.[6] For these reasons and in the light of what follows, it seems to be the only realistically feasible pricing policy compared to the various alternatives considered.

Legality of Basing Points. A basing point system may violate the antitrust laws in any of several ways: (1) as a *conspiracy in restraint of trade* if based upon collusion or conscious parallel action (Sherman Act, Sec. 1); (2) as an *attempt to monopolize* if imposed upon an industry by a dominant firm (Sherman Act, Sec. 2); (3) as an *unfair method of competition* if it avoids competition by adherence to a common course of action (Federal Trade Commission Act, Sec. 5); and (4) as *injurious to competition* if this is shown to be the effect of geographic price discrimination (Clayton Act, Sec. 2, as amended by Robinson-Patman Act). Although a basing point system has existed for decades in a variety of industries, it was not until the end of World War II that a new trend developed in the thinking of the courts with respect to this and related matters, as evidenced by several important cases.

The first set of cases to reach the Supreme Court in the postwar period involved the Corn Products Refining Co. (324 U.S. 726) and the A. E. Staley Manufacturing Co. (324 U.S. 746), both decided on the same day in 1945, and commonly referred to as the "glucose cases."

Both companies were engaged in the sale of glucose to candy manufacturers. The Court held that the companies' adherence to a single basing point system was in violation of the law; that this pricing method, which resulted in freight absorption on some sales and phantom freight on others, was discriminatory between customers and injurious to competition; and that the Staley company's defense that its prices were made "in good faith to meet the equally low prices of a competitor," was not acceptable since such prices had been quoted systematically. In subsequent decisions, handed down by a Court of Appeals in 1945 and 1946, the FTC was upheld in orders issued, involving not only single basing point systems, but also plenary systems (i.e., where each producing point is a basing point) and zone systems. In each of these cases, involving the sale of malt (152 F. 2d 161 [1945]), milk and ice cream cans (152 F. 2d 478 [1946]), and crepe paper (156 F. 2d 899 [1946]), the Court held that the characteristics of delivered pricing systems are such as to infer that there is agreement to avoid competition.

The second set of cases came in 1948. These involved the Cement Institute case, decided by the Supreme Court (333 U.S. 683) and the rigid steel conduit case, decided by a Court of Appeals (168 F. 2d 157). In the cement case, the Court upheld the Commission, declaring that the collective adherence by competitors to a multiple basing point system was "an unfair method of competition prohibited by the Federal Trade Commission Act" (p. 720); that such a practice was injurious to competition (p. 724); and that the good faith defense is unacceptable when the evidence reveals that price matching is consistent rather than sporadic for the purpose of meeting individually competitive situations (p. 725). And in the rigid steel conduit case, involving the sale of pipe shielding for electric wiring, the Court of Appeals upheld the Commission, finding strong evidence of agreement and declaring that the basing point system as such might be regarded as an unfair method of competition.

---

[6] *Cf.* Wilcox, pp. 221 ff.

Conclusion. These decisions led eventually to a settlement in steel. The industry shortly thereafter accepted an order agreeing not to participate in any pricing practices of a formula nature "which produces identical price quotations or prices or delivered costs," although delivered pricing or freight absorption is specifically permitted by the order "when innocently and independently pursued . . . with the result of promoting competition" (FTC Order 5508, issued August 16, 1951). Following the momentous cement and related decisions by the Supreme Court, a storm of litigation arose which lasted for several years. Articles, editorials, and books were written, hearings were held, bills were proposed, and legislation was enacted, all of which has been well accounted for elsewhere.[7] The outcome of all this may be summarized as follows. (1) The Commission's policy now is to accept the good faith defense as absolute, and its attention is focused on cases where *probable*, rather than mere possible, injury to competition is evidenced by illegal practices. (2) The courts will rely more on the Commission's own factual interpretation of the situation. (3) The Commission, in repeated assertations, has stated that it has never acted to prohibit delivered pricing or freight absorption when such practices were independently pursued with the result of promoting competition, nor does it intend to do so.

In other words, a policy of f.o.b. pricing with sporadic freight absorption has come to replace basing point systems in cement and several other industries. Although discrimination is still practiced, it is not systematic; the all-rail freight charge and common rate books have been abandoned, as have phantom freight and non-basing points. As long as demand remains strong, a policy of f.o.b. mill pricing is easily followed. But when demand slackens and idle capacity occurs, freight absorption is resumed. Whether such absorption will become so general as to result in a systematic matching of prices at each delivery point is something that remains to be seen. If a sufficiently high level of demand can be maintained, it is possible that the problem may never again emerge.

Distribution

It is evident by now that, in the area of distribution, the administration of the antitrust laws has not been with the aim of enforcing competition or preventing monopoly. The objective has been to protect the independent against the competition of the chains. The form which the attack has taken has centered around the advantages of size: the legal issue has been whether these advantages are great enough to restrict competitive opportunities sufficiently to constitute restraint of trade; the public policy issue has been whether the small competitor should be protected at the risk of impairing vigorous competition, or whether competition should be preserved at the risk of impairing the small competitor. The A & P case is an illustration in point.

---

[7] See Earl Latham, *The Group Basis of Politics*. For a fascinatingly written short account, see Wilcox, pp. 230–35.

The A & P Case. In a civil suit brought against A & P in 1949, the government charged three classes of violations: (1) illegal sales practices, (2) illegal buying practices, and (3) vertical integration.

Illegal Sales Practices. It was charged that the company engaged in local price cutting and selling below cost with the intent of eliminating competitors, while setting higher prices in less competitive areas to offset losses. The company, said the government, should more appropriately have followed a cost-plus pricing procedure. No evidence was shown, however, that the company increased its market share by using geographic price discrimination; nor was it shown that A & P recouped its losses by charging higher prices elsewhere, for in every market where the company operated it was also faced with competitors. In effect, the company was criticized for placing greater emphasis on demand elasticity than on costs in its pricing policy. Actually, what the company did was more a reflection of sound economic thinking than of unfair business behavior. Further, it operated on a low margin-high turnover basis, and this is a practice to be desired from normal competitive enterprise.[8]

Illegal Buying Practices. It was charged that A & P used coercive tactics to secure preferential treatment from suppliers. But what were these tactics? The company announced that it would buy only direct from suppliers and not through brokers, and that it would manufacture for itself if suppliers did not accept its terms. Accordingly, A & P was able to secure a broker's commission for performing broker's services; it received promotional allowances for advertising the products it handled; and it received discounts for the services it rendered producers. In other words, here was a company that, because of its oligopolistic position, was able to extract from other oligopolists certain concessions which it passed on to consumers in the form of lower prices. Certainly, consumers were not unhappy. But could it be argued that A & P exercised a monopsonistic (buyer's monopoly) position, leaving suppliers with no other alternative but to deal? Hardly. The evidence showed that the company purchased only about 10 per cent of the foodstuffs sold in national markets, and about 20 per cent of those sold in regional markets. Suppliers, therefore, if they were displeased with A & P's offers, could certainly have dealt with others. It seems that the government's charge of coercion should more correctly have been called successful bargaining.

---

[8] Concerning A & P's price policy, readers may recall the full-page advertisements that appeared in over 2,000 newspapers throughout the country. In reply to the government's charge that "defendants have regularly undersold . . . competing retailers . . . ," the company's advertisements read "To this charge we plead *guilty*. We confess that for the past 90 years we have constantly stepped up the efficiency of our operations in order to give our customers more and more good food for their money" Actually, what the company did should have been condoned, not condemned. It *made* competition, not hindered it.

Vertical Integration. The final charge leveled by the government was twofold in nature.

1. It held that A & P offset the low profits or even losses of its discriminatory retail operations with the profits of its manufacturing subsidiary. This it accomplished by having its factories charge higher prices to competitors than it charged its own stores. The meaningfulness of this charge may be questioned on at least three counts. First, many integrated firms subsidize the losses of one subsidiary with the profits of another (and hence violate the antitrust laws?). Second, A & P was evidently justified in doing so, for it realized definite savings by being able to consolidate shipments and by not having to incur the costs of soliciting business and transferring ownership. And third, the charge was absurd from a technical standpoint because, accounting-wise, the company could just as well have recorded lower transfer prices in its factory accounts thereby showing smaller profits or even losses in that subsidiary, and consequently higher profits in its retailing operations.

2. The government also claimed that the company's main central purchasing agency, Atlantic Commission Company (ACCO), which served as a produce broker for A & P as well as other distributors, had abused its dual function purpose. ACCO, it was charged, had sought to obtain produce at lower prices for A & P than it did for other distributors, and had attempted to establish exclusive dealings with suppliers and jobbers for the purpose of cutting rivals off from their sources of supply. Whether it could be clearly inferred that ACCO abused its dual function in favor of A & P is doubtful, for the evidence revealed that ACCO never attained a position even approaching that of monopoly. Suppliers and jobbers who were "victimized," therefore, could readily have taken their business elsewhere without being any the worse off.

The case was settled by the Consent Decree of January 19, 1954, the company being enjoined on several points: (1) In selling, A & P was forbidden to set low markups in particular stores with the intent of eliminating competitors by operating at a loss; but such intent would have to be proven, and could not be inferred from mere operation at a loss. (2) In buying, A & P was forbidden to exert pressure on suppliers that would prevent them from selling to competitors through brokers, offering discounts, or raising prices. (3) Concerning its vertically integrated structure, the company dissolved ACCO and was forbidden to buy food for, or sell food to, competitors, except that food processed in its own plants.[9]

---

[9] A fuller controversial treatment of the case appears in: M. A. Adelman, "The A & P Case: A Study in Applied Economic Theory," *Quarterly Journal of Economics*

Conclusion. The A & P case was not aimed at preventing monopoly or at enforcing competition; it was an attack by the government against the advantages of size. Certain relevant conditions that have come about with the growth of competing chains—a more competitive grocery industry, improved distribution methods at lower costs, better products at lower prices—in short, all the desirable consequences of competition, were hardly considered. The antitrusters, bent on preserving small business at the possible risk of impairing efficiency, were successful with the help of the court in imposing limitations on A & P that do not apply to its competitors such as Kroger, Food Fair, and others. This was accomplished despite the fact that the chains carry only a minor share of the business and that entry of new competitors with new methods is still occurring at a rapid rate. And it appears, in view of these happenings, that this pattern will continue in the future.

Resale Price Maintenance—Fair Trade

Resale price maintenance, popularly referred to as "fair trade," is a practice which permits the manufacturer or distributor of a branded product to set the minimum retail price at which that product can be sold. The purpose of this practice is to eliminate competition at the retail level in the prices charged for branded goods. Prior to 1931, attempts by manufacturers to establish resale price maintenance were repeatedly struck down as violations of the Sherman Act and the "unfair competition" provision of the Federal Trade Commission Act. In 1931, however, after a number of unsuccessful attempts to get a resale price maintenance bill through Congress, the retailers' associations, headed by the National Association of Retail Druggists, succeeded in persuading the California legislature to enact such a law.

Fair Trade in Action. The California law exempted from the state's antitrust act any resale price maintenance contract made between the seller and reseller of a branded product. In 1933, the law was amended to include the famous "nonsigner's clause": the terms of a resale price contract were made binding on *all* retailers if so much as a single retailer signed a contract. The amended California statute was quickly adopted by other states; by 1941, forty-five states (with the exception of Missouri, Texas, Vermont, and the District of Columbia) had resale price maintenance legislation on their books, under the euphemism of "fair-trade" laws. And high-pressure tactics were employed to enforce these vertical price agreements. Retailer associations blacklisted manufacturers who were unwilling to sign: lists were circulated to dealers disclosing which manufacturers had signed and which had not, from which retailers could decide which prod-

---

(1949), pp. 238–57; "The Great A & P Muddle," *Fortune* (December, 1949); and J. B. Dirlam and A. E. Kahn, "Antitrust Law and the Big Buyer: Another Look at the A & P Case," *Journal of Political Economy* (1952), pp. 118–32. For a shorter account, see Wilcox, pp. 406–11.

ucts to push and which to shelve or even boycott. And fair-trade committees of druggists were established which policed activities of other retailers, distributed contract prices, and in general threatened price cutters in various ways.

The state fair-trade laws were applicable only in intrastate commerce, i.e., when both parties to a contract were in the same state; in interstate commerce, as when the parties were in different states, the contracts were in violation of the antitrust laws. Since most branded goods moved between states, the federal antitrust laws would have to be amended if resale price maintenance were to be genuinely effective. This was accomplished in 1937 with the Miller-Tydings Act, which amended the Sherman Act. The Miller-Tydings Act provided that resale price maintenance contracts were exempt from the federal antitrust laws within those states where they were permitted by intrastate contracts. The antitrusters thus became limited in their prosecution of such contracts to only three states and the District of Columbia, while in the remaining forty-five states resale price contracts were given the go-ahead signal by Congress.[10]

Reaction. The validity of the Miller-Tydings amendment was not successfully contested until 1951, at which time the Supreme Court handed down its decision in the celebrated case of *Schwegmann Bros.* vs. *Calvert Distillers Corp.* (341 U.S. 384). Schwegmann Bros. had a New Orleans supermarket and the state of Louisiana had a fair-trade law with a nonsigner's clause. Calvert and Seagram had made resale price contracts on their whiskey with other retailers, but Schwegmann had signed no contract and proceeded to cut the price of "fifths." It reduced Calvert Reserve, for example, from the fixed price of $4.37 to the sale price of $3.25. The manufacturers sued and Schwegmann appealed. The Supreme Court, in a 6-to-3 decision, upheld Schwegmann, declaring that the Miller-Tydings Act: (1) did not give immunity to nonsigners, (2) that it contained no provision for nonsigners, and therefore (3) that it could not lend control to the prices of goods brought in from other states to be resold by nonsigners. In other words, the Miller-Tydings Act, in the opinion of the Court, was applicable only to signers, not to nonsigners, of resale price contracts, and any attempt to apply it to nonsigners was, in the opinion of Justice Douglas speaking for the Court, "price fixing by compulsion . . . [and] resort to coercion."

The result of this decision was sensational. With Schwegmann and the Supreme Court playing the role of innovators, a nationwide price war on fair-traded items went into effect in the true Schumpeterian manner.

---

[10] The Miller-Tydings Act was put through Congress as a rider to the District of Columbia Appropriations Act, just prior to Congress's adjournment. Though President Roosevelt signed the bill under protest, objecting both to the substance of the law and to its method of enactment, he had no other alternative: he had to accept the rider or else deprive the District government of its needed revenues for further activities.

Almost overnight, large and small stores in almost 50 American cities had cut prices on numerous items anywhere from 20 to 50 per cent, and on some items by even more.[11] Macy's and other large stores throughout the country published daily price lists in the newspapers, and many stores reported that the number of shoppers even exceeded that of the Christmas season.

The heyday lasted about five weeks, by which time retailers' stocks were depleted and manufacturers had refused to replenish them. With a deficit in supply, prices moved back to their previous levels. And then the pressure for correction came. Some 1,300 local, regional, and national trade associations bombarded Congress with letters, telegrams, phone calls, and delegations of visitors, demanding correction of the deficiencies in the Miller-Tydings Act. Despite strong opposition by labor, agriculture, the antitrust agencies, and consumer and other organizations, both houses of Congress passed, by substantial majorities, the McGuire-Keogh Fair Trade Enabling Act, and the bill was signed by President Truman on July 14, 1952.

Counterreaction. The McGuire Act put the law back to where it was prior to the Supreme Court's decision in the Schwegmann case. The Act, passed as an amendment to Section 5 of the Federal Trade Commission Act, took explicit recognition of the problem by extending the federal exemption from the antitrust laws to include nonsigners. The Act thus reversed the Schwegmann decision by allowing the enforcement of interstate contracts against all dealers in a state when such contracts have been signed by any one of them.

Again Schwegmann challenged. This time it sold insulin manufactured by Eli Lilly & Co. below the fixed resale price. Lilly sued, under the Louisiana law, and the state court granted an injunction. Schwegmann took the case to the federal Court of Appeals, contending that the nonsigner's clauses both in the state and federal laws were unconstitutional, but lost the case in 1953 (205 F. 2d 788). It then went to the United States Supreme Court, but in October, 1953, that body refused to hear the case (346 U.S. 856). The following year, the same court again refused to review the decisions of lower courts which had upheld fair-trade practices against the Sam Goody company, a New York phonograph dealer, and S. Klein, a New York and Newark department store. It appears, therefore, that by the Supreme Court's refusing to hear these cases, the McGuire Act is firmly established.

Conclusion: Future Outlook. Two questions still remain to be answered: (1) What is the future for resale price maintenance, and (2) what are its economic consequences?

With respect to the future, the issue is far from closed. By the end

---

[11] As an extreme example, Bayer aspirin was reduced from 59 cents to 4 cents, a cut of 93 per cent. Other reductions were not as severe, but discounts of 40 or 50 per cent on items were quite common.

of 1956, under continuous pressure by the opposition forces, four sta
had ruled that resale price maintenance laws are unconstitutional, twe
states had declared the nonsigner clause in such contracts to be unconsti
tional, and one state supreme court had held the law in general to be
operative. But this is just the beginning, and litigation is continuing to t
the validity of the laws. Recent evidence of weakness has appeared in s
eral instances. In the Wentling case decided in 1950 (185 F. 2d 903)
was held that a mail-order house, which was bound by resale price c
tracts between manufacturers and retailers in its own state, was not bo
by the nonsigner's clause when selling to other states. Sales by mail acr
state lines thus fall beyond the scope of the nonsigner's clause. It see
therefore, that mail-order houses in the District of Columbia or in st
without fair-trade laws would appear to be completely free of resale p
controls. And in two cases brought by General Electric involving p
cutting by retailers, one in 1951 and the other in 1955, the court h
that it is the responsibility of the manufacturer to police his own f
trade contracts and see to it that they are diligently enforced on a gen
scale. Most manufacturers have evidently found this too costly to do
witnessed by the rapid growth of discount houses which, to a large deg
deal in what is supposed to be fair-traded items. Perhaps these events
but straws in the wind. More probable, considering the mushroom-
growth of discount houses and the changing pattern of distribution t
has occurred in the past decade, it appears that resale price maintena
will decline in relative significance in the years to come.[12]

Concerning the consequences of fair trade, the issues have been
bated for years. Those in favor of it have argued that it: (1) protects c
sumers by preventing deception, (2) protects retailers, especially s
ones, by preventing price warfare, and (3) protects manufacturers
preventing price cutting to where it may eventually cause the loss
market.[13] The objective, in large part, is to prevent the use of loss lea
as a means of increasing sales. Those opposing fair trade have argued
it: (1) facilitates price rigidity and price agreements among manufac
ers and retailers dealing in competing products, (2) suppresses comp
tion in retailing by preventing lower prices even among the most effic
retailers, and (3) freezes distributive channels and brings about hig
costs by encouraging nonprice competition (e.g., increased advertis

---

[12] Thus in February, 1958, General Electric officially announced its aban
ment of fair trade. Other companies have followed a similar pattern in recent y

[13] The classic example is that of the Ingersoll watch, "the watch that made
dollar famous." Some retailers used the watch as a loss leader, the price event
being driven down to 57 cents. Most retailers were unable to take the loss requ
to sell the watch at this price, and consumers would not pay more than this am
Retailers thus had to drop the line and the manufacturer eventually lost the ma
(see *Congressional Record,* July 1, 1952, pp. 8935–36).

salesmanship, services) in place of price competition. The objective here appears to be that of replacing the rigid structure of prices and distribution caused by resale price maintenance with a more flexible structure that would emerge automatically if there were a free play of market forces.

Of what significance are these arguments? Actually, resale price maintenance is not as serious as the pros and cons would suggest, for at least two reasons: limited application and ease of evasion.

Limited Application. Resale price maintenance is most suitable for products that are easily identified (i.e., branded), widely used, frequently purchased, and moderately priced—and, particularly important, where the cost of raw material is a relatively small proportion of the final price, so that substantial fluctuations in materials costs are not significant enough to exert pressures for price changes. Thus, drugs, cosmetics, liquor, tobacco, appliances, sporting goods, and the like, have been susceptible to price maintenance, while clothing, furniture, jewelry, hardware, and foodstuffs have not. In the aggregate, less than 10 per cent of the value of all retail goods in the United States are sold under price maintenance conditions.

Ease of Evasion. Retailers can readily evade selling at maintained prices by employing a number of successful dodges. These include offering premiums, trading stamps, liberal trade-in allowances, gifts, employee discounts, bonuses, and special deals, and by conducting special sales on goods that have been slightly damaged (e.g., "accidentally" scratched in inconspicuous places) or slightly used (e.g., one-time "demonstrator" models). Thus, in view of the difficulty on the part of most producers in enforcing fair trade, plus the fact that the discount sellers buy in quantity and pay cash, it appears that manufacturers are coming increasingly to regard resale price maintenance as a thing of the past.

Summary

This section has surveyed the current status and future outlook of various uncertainty areas in the field of competition and its regulation. The present state of the law in these areas may now be summarized.

1. The courts have consistently struck down conspiracies among competitors in restraint of trade. Monopolization, on the other hand, has been treated less severely. It appears now that the power to abuse, even though lawfully acquired and never exercised, is sufficient to rate condemnation. In the case of single-firm monopoly, the courts have shown an increasing tendency to limit market powers and to reform market practices so as to encourage the entry and growth of new competitors. But with close-knit combinations the courts have been reluctant to order their break-up as long as other solutions appear possible. But the question still remains: what is monopoly power? The courts have employed market-

share figures as a criterion, but have failed to establish how much of a market share is legal and how much is not. This problem will be considered further in the next section.

2. Patent abuse and the power of patent monopoly has been significantly weakened in recent years. Cross-licensing and patent pooling, though not illegal as such, have usually been condemned by the courts when used as a means of eliminating competition. Remedies have included the requirement of royalty-free licensing and the provision of necessary know-how by the patent holder. It appears that the courts will continue to move in the direction of preventing the abuses of the patent grant. Similarly, with respect to trade-mark abuse, the courts have acted increasingly to prevent the use of trade-marks as a tool for promoting price discrimination, market exclusion, and market sharing in the international sphere (i.e., cartel arrangements) among competitors.

3. Exclusive contracts and dealings, though illegal under the Clayton Act if they tend substantially to lessen competition, have been treated inconsistently by the courts. Accordingly, the apparent policy of the Federal Trade Commission is to confine its orders to where it can show a probability of substantial injury to competition in a market as a whole. The question remains of course: when is competition *substantially* lessened? The ultimate decision lies with the courts, and no clear-cut pattern seems to exist on the basis of which one can establish objective guides.

4. Price discrimination, under Section 2 of the Clayton Act, is unlawful "where the effect of such discrimination may be to substantially lessen competition or tend to create a monopoly *in any line of commerce*." Exceptions are provided however, for price differences based on: (*a*) grade, quality, or quantity sold, (*b*) differences in selling or transportation costs, or (*c*) good faith to meet competition. The seller charged with illegal discrimination thus has two defenses available: one is the cost defense, meaning that his price differences were justified by differences in cost; the other is the good faith defense, meaning that his lower price "was made in good faith to meet the equally low price of a competitor." Both defenses have been seriously weakened by the Federal Trade Commission, however, and it seems that the courts will uphold the Commission where it appears that the defendant has *constantly* maintained a discriminatory pricing structure (as in the basing point system in the Cement case) or a regular schedule of quantity discounts (as in the Minneapolis-Honeywell case).

Briefly, the legal status of discounts is as follows. Brokerage payments are illegal where no brokerage function is involved, even if the benefits are passed on in lower prices to consumers. Allowances and services must be given "on proportionally equal terms," but the Commission has established no clear rule as to whether proportionality should be based on dollar sales volume, on the cost to the buyer of the seller's services rendered, or on the value of the services to the seller. Variations of all three have been used and held legal. And quantity discounts may or may not be legal, even if justified by costs, depending on whether the Commission feels that they are injurious to competition (as in the U.S. Rubber case in 1939).

5. Delivered pricing, on the basis of the glucose cases, the cement case, and the rigid steel conduit case, has been held illegal where basing point systems are employed, where the term "basing point system" refers to a formula method of pricing that produces identical prices among competitors. Where a pricing system prevails that produces a central tendency toward identical prices, but from which there are frequent variations even though something approaching a basing point system is employed (e.g., sugar industry), this is not illegal. According to the Commission, it is a "planned common course of action, understanding, or agreement" that is unlawful. Delivered pricing or freight absorption "when innocently and independently pursued, regularly or otherwise, with the result of promoting competition" is not illegal, nor does the identity of delivered prices at any destination, or conscious parallel action, necessarily prove violation of the law or evidence of conspiracy. Since 1953, the Commission, having a Republican majority, has adopted the policy that probability and not mere possibility of injury to competition must be evidenced, and that the good faith defense must be absolute rather than procedural. It appears less likely, therefore, that the Commission will prosecute delivered pricing systems in the future as vigorously as it has in the past.

6. In the field of distribution, the prevailing philosophy has been to protect independents from the competition of the chains, and it appears that this philosophy will continue in the future despite the misnomer that this is "maintaining competition." The chains have brought the consumer, through improved distribution methods, a greater diversity of products at lower prices, just as is to be expected of competition. It seems, therefore, that if a policy aimed at preserving smallness and independence is to be followed, it ought to be stated as such and not be masqueraded under the title of competition, so that consumers may become aware of the price to be paid. Smallness and independence may well be desirable institutions worth preserving, but the reasons are probably other than economic and should be recognized as such.

7. Finally, as to resale price maintenance, the indications now are that it may, in large part, soon be an institution of the past if it is not already. As the courts place increasing emphasis on self-regulation by manufacturers, the latter in turn find it unprofitable and difficult to enforce contracts. On the whole, it seems that manufacturers have suffered no loss as a consequence, and have probably even profited from increased sales.

MEASUREMENT OF ECONOMIC CONCENTRATION

The growth and importance of big business in the United States has resulted in charges, frequently made and widely believed, that: (1) the concentration of economic power is centered in the hands of a few corporate giants, (2) that this concentration has grown over the years, and therefore (3) that there has been a general "decline of competition." Upon close examination, it appears that of these three charges, the evidence shows the first to be only somewhat true, and the second and third to be entirely unfounded. Let us see why.

Concentration Data

A substantial number of studies have been made since the early 'thirties for the purpose of discovering the degree of concentration of assets, employment, income, and sales in the hands of a few large firms. The groups studied have included nonbanking corporations, manufacturing as a whole, particular manufacturing industries, and the output of manufactured products. Some of the typical findings may be cited.

In his article "The Measurement of Industrial Concentration," in the *Review of Economics and Statistics* (1951, pp. 275–77), M. A. Adelman cites these data. In 1947, the 200 largest employers in the nation accounted for almost 20 per cent of total employment in private nonagricultural establishments, and the 200 largest corporations held 40 per cent of all corporate assets and between 20 and 25 per cent of all income-yielding wealth. In the same year, 163 manufacturing firms with more than 10,000 employees accounted for 30 per cent of the employment, and 133 firms with assets over $100 million held 40 per cent of all the assets.

The National Resources Committee in its *Structure of the American Economy* (pp. 240–58) states the following information. In 1935, among 275 manufacturing industries, 8 firms hired more than half of the workers in each of 131, and 4 firms hired more than half of the workers in each of 75. And in 1947, the 4 largest firms accounted for more than half the output in 150 out of 452 such industries, and for more than 50 per cent of the output in 46 of the industries.

Finally, W. L. Thorp and W. F. Crowder, in their *Structure of Industry* (TNEC Monograph No. 27, 1941, Part III) point out that among 1,807 products reported in 1937 for the *Census of Manufacturers*, the four largest producers accounted for over 85 per cent of the output in one fourth of the cases, for over 70 per cent in nearly half the cases, and for nearly 50 per cent in three fourths of the cases.

Evaluation

The above type of evidence is typical of the "facts" usually cited to support the contention that there has been a decline of competition. At first glance the evidence seems conclusive, but upon closer analysis it loses much of its impressive nature, once it becomes evident that the results will vary depending on the way the calculations are made.

Choice of Base. The degree of concentration will vary depending on the base chosen, such as all businesses, all manufacturing, all corporations, all nonfinancial corporations, all industries, and so on. The concentration figures cited above for all nonbanking corporations include, for instance, railroads and utilities whose monopolistic powers are regulated by public agencies and whose assets amount to half the assets of the 200 largest corporations; they also include several firms operating in highly competitive industries (e.g., A & P, Macy's, Sears Roebuck) as well as other firms that do not necessarily exercise significant control over their output and input markets. The figures, therefore, do not indicate the real degree of unregulated monopoly.

Choice of Unit. The degree of concentration will vary depending on the unit of measurement chosen, such as a plant or a firm, or a single-product or multiple-product firm. The figures above apply to a heterogeneous conglomeration of industries some of which are highly competitive, some that are moderately so, and some that are virtually monopolized. Of the 450 manufacturing industries considered, only four contain two thirds of the concentrated assets in the group. Further, the ratios are obscured because they pertain to the three, four, six, or eight largest firms in an industry, without revealing the degree of domination by a single firm. The ratios thus disclose little as to the extent of competition or monopoly.

Choice of Index. Still another factor affecting the measure of concentration is the index of concentration used, such as assets, employment, output, income, or sales. The concentration ratios by industries are normally based on the *Census of Manufactures*' classification, in which industries are defined, in part, by the materials and processes they employ. Accordingly, some firms producing noncompeting products are grouped in one industry, while others producing competing products are grouped in separate industries. Similarly, the ratios for particular goods are based upon a classification that defines products, in part, by the materials, fabrication processes, and degree of manufacturing integration involved in their production, so that the result may be multiple listings for a single commodity. In some cases, the concentration ratio is seriously understated because the output figures are national and markets are regional, or because heterogeneous goods are lumped together into a single category. In some cases the ratios are greatly overstated because the figures are limited to domestic production with competition from imports ignored, and because readily substitutable products are listed in unrelated categories. Thus the data on concentration reveal little either as to the structure of markets for particular goods or as to the index of concentrated power.

It is apparent, therefore, that the measures of economic concentration are not measures of monopoly power as is often contended. At best, these measures may reveal the results of monopolistic restriction or collusion, or of innovation, market development, and lower costs and prices. But they may also conceal the influence of potential competition, and the existence—on the other side of the market—of countervailing power.[14]

Conclusion

If the concentration data cited above do not really prove that the "engine of monopoly power" has overtaken the American economy, what then can be said about the extent of monopoly in the United States today? The results of four recent basic studies, done by Nutter, Adelman, Kaplan, and Stigler, provide the answer.[15]

G. Warren Nutter defined as monopolistic all industries in which the four largest firms produced more than half the output, and then measured the extent of monopoly according to the share of national income originating in these industries. Comparing the extent of monopoly in 1899 and 1939, he found that there was: (1) a relative increase in monopoly in financial enterprises, (2) no change in agriculture, services, trade, construction, and public utilities, and (3) a decline in manufacturing, mining, transportation, and communication.

M. A. Adelman, continuing Nutter's work, compared the extent of concentration for the years 1901 and 1947 in manufacturing industries. His conclusions were that: (1) "the odds are better than even that there has actually been some decline in concentration," (2) it is "a good bet that there has at least been no actual increase," and (3) that for the economy as a whole, "the extent of concentration shows no tendency to grow, and it may possibly be declining."

A. D. H. Kaplan, in a study for the Brookings Institution, found that: (1) the number of business firms per thousand of population increased slightly when 1900 is compared with 1950 and 1929 with 1949, (2) big business has grown at a rate proportional to the economy as a whole, and (3) that the 100 largest industrial corporations is not a static list, but rather a dynamic one in which new firms are always emerging and old ones declining.[16]

---

[14] For a fuller evaluation of concentration measures, see: G. W. Nutter, *The Extent of Enterprise Monopoly in the United States*, pp. 11–19; Wilcox, pp. 835–36, and his "On the Alleged Ubiquity of Oligopoly," *American Economic Review* (1950), pp. 67–73. The expression "countervailing power" has its own profound implications, as expressed in J. K. Galbraith's well-known study, *American Capitalism* —a book that should be on the reading list of every executive and policy maker.

[15] G. W. Nutter, *op. cit.* (1951), pp. 35–43; M. A. Adelman, "The Measurement of Industrial Concentration," *Review of Economics and Statistics* (1951), pp. 293–95; A. D. H. Kaplan, *Big Enterprise in a Competitive System* (1954), pp. 71–73, chaps. VI and VII; G. Stigler, *Five Lectures on Economic Problems* (1949).

[16] For an excellent and lively summary of this outstanding study, see A. D. H. Kaplan and A. E. Kahn, "Big Business in a Competitive Society," *Fortune* (February, 1953).

Finally, George Stigler, in a less extensive study than the previous ones, estimated that: (1) about one third of the income produced in 1939 originated in industries characterized by individual monopoly and compulsory cartelization, and (2) about two thirds came from industries that were competitive. Stigler's finding is not conclusive, but it accords fairly well with the observations made by the previously mentioned authors.[17]

These studies go a long way in challenging many accepted beliefs and allegations of certain laymen, economic reformers, political propagandists, and professional economists who contend that there is a relative or even complete absence of competition in big business. The conclusions and supporting data indicate that effective competition can and does exist in large-scale enterprise, and that the traditional assumptions of nineteenth century (pure) competition are unrealistic in our twentieth century economy, characterized by rapid product development, market growth, and vast technological advancement. It is only in the manufacturing sector of our economy that the problem of unregulated monopoly exists, and even here the scope is narrowed down to a select group of industries where concentration is high. Some of these include, in the industrial field, petroleum, rubber, glass, primary metals, newsprint, and heavy equipment; in consumer durables, autos, radios, sewing and washing machines, refrigerators, and vacuum cleaners; and in lighter consumer products, cigarettes, soap, matches, and light bulbs. The tendency of economic reformers to identify the major producers in these fields as monopolies merely because of their size, only serves to distort the real nature of the problem. Production in these industries is characterized by a small number of large firms, so that the antitrust problem is one of oligopoly, not monopoly. And economic theory does not say that oligopoly is not fiercely competitive; it only states that there may be a stronger tendency to avoid price competition.

BIBLIOGRAPHICAL NOTE

A recent and probably the most comprehensive treatise analyzing the economic relations between government and business is Clair Wilcox's, *Public Policies Toward Business*. More than just a text, it represents a genuine contribution to the field, reflecting Professor Wilcox's mature thought expressed in a sophisticated manner and interesting style. A few of the many other works in the field that are worth careful reading are: on the objective of antitrust policy, C. Edwards, *Maintaining Competition*, chap. 1, and A. G. Papendreou and J. T. Wheeler, *Competition and Its Regulation*, chap. XIII; on the significance of major court decisions since the World War II period, J. B. Dirlam and A. E. Kahn, *Fair Competition: The Law and Economics of Antitrust Policy*, chaps. III and V, and Papendreou and Wheeler, chaps. 15–18, which also contains lengthy excerpts from leading decisions. The various techniques of measuring monopoly power, which have been debated by economists for years, has been the subject of a number of journal articles. For a survey and evaluation of the proposed measures, see F. Machlup, *The Political Economy of Monopoly*, chap. 12; also, a pamphlet by the National Industrial Conference Board containing opinions of various economists, *Economic Concentration Measures, Uses and Abuses* (Studies in Business Economics No. 57). Finally, in addition to Wilcox's and the other works mentioned above and in the footnotes, two general books providing a fuller survey of many of the topics treated in this chapter are G. Stocking and M. Watkins, *Monopoly and Free Enterprise*, and V. Mund, *Government and Business*, 2d ed. The latter contains an appendix with extracts from landmark cases, while a fuller treatment may be found in a handy little book by I. Stelzer, *Selected Antitrust Cases.*

---

[17] For a short extract of the highlights of Stigler's study, see A. D. Gayer, C. L. Harriss, and M. H. Spencer, *Basic Economics*.

QUESTIONS

1. (*a*) What are the antitrust laws? (*b*) List the chief ones and the years in which they were passed.
2. Outline briefly the nature of the antitrust laws enforcement process.
3. What, briefly, are the various practices held to be illegal under the antitrust laws?
4. Explain the meaning of a restrictive agreement. What is likely to be the opinion of the courts on such matters? Support your answer.
5. What is the current trend of the courts with respect to: (*a*) monopoly *per se*, (*b*) vertical integration, and (*c*) mergers?
6. State what is likely to be the decision of the courts with respect to each of the following: (*a*) The decision of a patentee to withhold an important, even life-saving invention, from general use. (*b*) The policy of a patent holder of requiring that a licensee purchase other products as a condition for licensing a patented product. (*c*) The attempt of a patent holder to fix the price charged by a licensee; by several licensees. (*d*) Cross-licensing and patent pooling.
7. Explain how trade-marks can and have been employed by various firms as devices for enhancing long-run profits.
8. What seems to be the current position of the courts with respect to exclusive dealerships?
9. With respect to the Robinson-Patman Act on price discrimination, explain: (*a*) the test of illegality, and (*b*) the defenses available to the violator in such cases.
10. (*a*) In the light of your answer to question 9, evaluate the significance of the Robinson-Patman Act with respect to its enforcement, i.e., what enforcement difficulties does it involve, especially from an economic standpoint? (*b*) On balance, what do you think are the net consequences of the Act?
11. Outline the position of the courts or FTC concerning: (*a*) brokerage and allowances, (*b*) quantity discounts, and (*c*) functional discounts. Cite major cases and discuss.
12. What recommendations do you suggest for improving the effectiveness of the Robinson-Patman Act?

13. (*a*) Explain the meaning of delivered pricing. (*b*) Why might delivered pricing be discriminatory? (*c*) What chief factors will determine the significance of delivered pricing in a discriminatory sense?
14. What is the current attitude of the courts and of the Federal Trade Commission concerning delivered pricing? Discuss, in the light of the major cases.
15. Explain what you believe to be the underlying philosophy of the government in the area of distribution.
16. Has there been a "decline of competition" in the United States? Discuss.
17. "The antitrust problem today is not monopoly, but oligopoly." Explain.

Chapter 10

CAPITAL MANAGEMENT

ADMINISTRATIVE ASPECTS OF CAPITAL MANAGEMENT

Function of Top Management

Of the various types and classes of business problems, the most complex and troublesome for the decision maker are likely to be those relating to the firm's capital investments. These problems are typically too serious to be decided by any group below the top management level, and are often so complex as to require extensive applications of time and labor in disposing of them. Such attributes of capital investment problems stem from a host of factors which influence the direction the decision makers will take and upon which the decisions will significantly impinge.

Relatively large sums of money are typically involved in the budgeting of capital investments. Whether or not these funds are internally available, their commitment is a matter for top-level decision. Where resort to outside sources is necessary, negotiations with commercial and/or investment bankers will be involved in which the firm's future income and resources will in some way be committed to satisfying the claims of the capital contributors. If internally available funds are adequate, their commitment will at least affect the firm's working capital position and liquidity, will introduce problems of resource deployment, and will probably require a review of dividend policy.

Long-Run Nature of Capital Decisions. The area of investment decision making is further complicated and enhanced by the "long-run" nature of the fund commitments. The investments become "sunk," and mistakes, rather than being readily rectified, must often be lived with until the funds can be withdrawn through depreciation charges or (distress) liquidation sale. Obviously, investment problems are not all equally important, their significance varying largely with the scope of the programs contemplated. But it is clear enough that the manner in which these problems are disposed of will at least color the corporate life of the enterprise and, in particular circumstances, might well seal its doom if improvidently handled.

These considerations tend, in many cases, to cause managements to relate capital expenditures to the long-run position of the firm in its economic environment (although, as we shall see, many businessmen require a rather quick—usually three years—return of their capital).[1]

Review of Investment Alternatives. Such grave consequences, therefore, call for studious examination of investment programs by top management and its staff. But even abstracting from these dire possibilities, it is incumbent upon the management of a profit-seeking firm to evaluate many possible investment alternatives and to select rationally from among them, after careful deliberation, in committing the limited resources under its control. Cost studies and market analyses will often be required, and cash flow projections made in which cash expenditures are related to cash receipts. The business decision maker receives some of his greatest challenges in this area, and the quality of the decisions depends not only on his own abilities and judgment, but on the abilities of the staff on whom he must rely for the studies and reports channeled to him, and on the administrative organization that has been established to serve as the framework within which such functional problems can be handled.

Need for a Capital Budget. The tool employed for these purposes is known as the *capital budget*, and it encompasses all proposals involving expenditures for capital purposes. Because such expenditures lead to the acquisition of long-lived assets, a "planning horizon" extending beyond the current year (typically about three years, rarely more than five) is involved, though periodic (probably annual) revisions of the plan will be effected. Implicit in such an approach is the making of a (long-range) forecast. Fraught as this is with the many dangers previously discussed, it is a necessary forerunner (whether performed crudely by some rule-of-thumb method or scientifically by the employment of econometric techniques) to the construction of the capital budget. Where the expenditures are designed to reduce costs without affecting the revenue-producing ability of the firm, or where the revenue function is relatively stable and not particularly sensitive to moderate changes in the economic environment, e.g., electric power demand, the forecasting problem is greatly simplified. But where new product lines are being added, new markets invaded, or the expansion of existing products is contemplated, the forecasting problem is compounded by numerous uncertainty factors many of which may be created by the firm's very act of aggression. This is one of the economic facts of life and, rather than try to escape it by ignoring its existence, the progressive firm seeks to live with it and to overcome the obstacles it presents by employing a modern kit of forecasting tools. A margin of error must, however, also be planned for, and this is accomplished by building flexibility into the capital budget.

---

[1] There is an apparent inconsistency here between the businessman's insistence on short payoffs, and his couching of capital planning within a long-run environment. Partly this is due to the fact that the "long run" for many businessmen is not a secular type of long run with which the time-series analyst is usually concerned; partly it is due to the difference in the businessman's mind between the longer useful life of the plant he constructs and the much shorter period in which he requires his investment be returned.

Building the Capital Budget

Flow of Proposals. The elements which become a part of the finalized capital budget all begin with an idea on the part of someone, from the president on down to the shipping clerk. In the process of its conversion into a part of the budget, a proposal will go through a screening process the purpose of which is to determine its merits and faults. The author of a proposal may often be so entranced by its possibilities that he will overlook an obvious weakness which makes it quite unrealistic. The screening procedure is thus a weeding-out process, and those proposals which emerge from the first screening are usually ready to be quantified in terms of monetary outlays, timing, and benefits expected.

It is impossible to generalize very much as to the specific procedures employed by firms in constructing their capital budgets, because such a great deal of diversity is possible and does exist. But this does seem to be generally the case (as one might logically expect): that detailed project proposals relating to cost reduction, replacement economies, and others of a technological nature more often originate from operating personnel and engineers. Budgets calling for aggressive expansion, or invasion of new markets, are more likely to reflect top management's initiative.

Limits of Discretionary Spending. In the typical capital appropriations control program, the screening process is a multilevel affair progressing ultimately through the board of directors, although it is quite common for a certain amount of autonomy to be granted to the lower managerial levels in carrying out expenditures within preassigned limits without the need for securing approval from above. Should a given expenditure exceed the preassigned discretionary limits, the proposal will lend itself to ready processing if it affects only one department or plant and therefore requires no consultation with other plant or department managers, although approval from above would still be necessary to assure the allocation of funds for the project.

The Hierarchy in the Budget Approval Process

The Plant Manager. In the channeling of a given proposal up the ladder of authority, the first close official look will be given by the plant manager or department head, who, if he approves it, will submit it to his superior, and so on up to the top management level. The number of levels through which a proposal will have to progress when initiated by someone near the bottom rung, e.g., one of the production engineers, will vary from firm to firm depending on the organization's size and complexity.

The Divisional Head. The plant manager's budget estimates will be assembled by him, with the aid of his department heads and selected members of his staff possessing the proper technical training. A careful check of the data is mandatory because the manager's approval is equivalent to his adopting the project proposal as his own. Where a divisional head is interposed between the plant manager and the executive committee his approval will also be required.

The Budget Director or Budget Committee. Meanwhile, a similar flow of requests for capital funds will be taking place in the other divisions of the firm, all channeled through the divisional heads toward one focal point—the budget director, or officer acting in the equivalent capacity. This officer, by himself or in committee, will screen and consolidate all requests into a single, well-ordered, meaningful statement in which the various projects will probably be ranked in order of their estimated profitability. At this point the budget is presented to the executive committee (comprising the president, the chief financial officer, and the other general or executive vice presidents) or to a purposefully constituted budget committee (made up of much the same, though perhaps fewer, officers). This is the final review given by the company's operations officers, and after this group approves it the budget is submitted to the board of directors.

The Board of Directors. The board's concern with the capital budget is not with its minute details, but with its over-all objectives. The directors will want to know what areas of the company's operations are to receive major emphasis in the expansion, the sources of the funds to be used to finance the various projects, the estimated productivity of these proposals, the total size of the expenditures contemplated and whether this is expected to have any effect on current or near-term dividend-paying ability, and the consistency of the over-all plan so as to avoid the creation of imbalance in the productive capacities of the elemental parts of the firm.

The budget might well be reshaped at this point, to whatever degree, as a reflection of the board's concern over its size or timing, or because the directors might take issue with the emphasis on certain aspects of the company's operations. However, most of the controversial aspects of the proposed budget will, by this time, have been anticipated and fully debated by the executive officers, so that it is rather unusual for the board to reject a plan brought formally before it. Modifications, if any, will be ordered and, subject to these changes, the budget will be formally adopted, but responsibility for carrying out the programs specified in the capital budget is turned over to the president as the chief operating officer.

Expenditure Authorization

The board's acceptance of the capital budget represents simply its acceptance of an expectational plan for the ensuing year, but does not provide anyone with specific authority to start spending money. This authority may actually not be obtained for a particular project until quite some time after the board's approval of the capital budget. The exact way in which the specific authority is secured is determined by the operating procedures set up within the firm and will therefore vary from one company to another. In the typical case, each project will be examined and analyzed again by the respective plant managers or department heads who will have one of their staff members (probably the project author) prepare a final, up-to-date, detailed report, taking into consideration all conditions that might have changed in the meantime. This report, and the data going into it, will serve as the official basis for the request for authorization of expenditure of funds, and will be tested later against the actual performance of the project when it is in operation. Depending on the company, the application for funds will be submitted over the signature of the plant manager or department head, as the case might be, "through channels," until it is finally approved by the highest responsible officer. This might be the controller, the treasurer, the executive vice president, the chairman of the finance committee, or even the president himself. The expenditure having been authorized, the rest is a matter of bookkeeping and controls exercised by the accounting department which will recognize the authorization by placing the project on its books and assigning to it its own account number.

This issuance of a "birth certificate" for the "baby" is explicit recognition of its existence, but whether it will prove to be lusty and vigorous and contribute to the general well-being of the "family" is, of course, uncertain. Much will depend on factors outside of management's control—an economic recession, unexpected price competition in the industry, a sudden development of consumer resistance, the appearance of competing products serving the same end use—but a great deal will also depend on management's astuteness both in the original choice of the project from among the alternatives available and in their direction of it during its productive life.

FORWARD PLANNING OF CAPITAL EXPENDITURES

The diversity of, and resultant impossibility of generalizing about, capital planning among business firms has already been commented upon. Some findings concerning the behavior of particular firms under certain circumstances are, nevertheless, worth commenting upon both as an interesting observation of the particular firm in question and as a probable indication of the manner in which many firms in similar circumstances might be expected to react.

Capital planning is very much a reflection of management philosophy—an ephemeral quality to try to generalize about. Much of management's attitudes depends on the nature of the business—companies in fast-growing fields are likely to be aggressively expansionist; stagnant industries are more than likely to be comprised of sleepy, slow-moving enterprises. But within an industry, considerable variation among firms exists. For example, Federated Department Stores has proven to be an aggressive, consistent growth company in a field which has long been considered mature.

An Expansionist Firm in a Growth Industry

In the chemical industry—a prime growth area—Air Reduction provides an example of a management determined to get more than its existing share of the industry's growth by pursuing a policy in which:

> . . . production shall push sales. By this is meant that at all times the capacity of our production and distribution facilities shall be in excess of current sales so that our sales force can continue to sell and no customer will be turned away. In order to implement this policy we have . . . expended large sums for new and improved facilities . . .[2]

Such a policy involves a very long-range view, an attitude which many firms are wont to disclaim. Again, in reference to Air Reduction:

> In its forward planning your management emphasizes the long range point of view. We believe that our country and its economy will continue its dynamic growth, and it is our firm objective that your Company shall always be prepared to serve the growing needs of the American consumer. We recognize that the industrial and commercial expansion ahead of us will not be without its peaks and valleys. We are confident, however, that with the continuing upward trend of economic growth the peaks of business activity will reach progressively higher levels.[3]

Example of a "Mature" Industry

While expansion and growth seem to be the major factor underlying capital expenditure planning, other (and related) considerations frequently play important roles. The steel industry is an interesting case in point. In an attempt to escape the feast-and-famine, prince-and-pauper forces to which they were long subject, the major steel firms took giant steps toward market and product diversification, thus freeing themselves of their dependence on heavy capital goods demand. In doing so, the industry moved its products closer to the final consumer. Where the rail-

[2] *Air Reduction*, Annual Report of 1953, p. 14.

[3] *Air Reduction*, Annual Report of 1955, pp. 12–13.

road industry was the single, most important steel consumer for decades, it is today the sixth largest user of steel. The automobile industry takes about 20 per cent of our nation's steel,[4] with construction, machinery, tin cans and containers, appliances, and other consumer products ranking among the top users. Approximately two thirds of total steel output is today of the lighter grades—wire, tubing, sheets, etc.—reflecting the rising importance of the automobile, container, appliances, and other light-steel consuming groups.

The above-described change in industry character might be argued to have been induced by forces exogenous to the industry itself. Yet those companies in the industry which responded successfully to the technological changes in the economy should be credited with aggressively developing and promoting steel products for uses which might otherwise have been supplied by other metals. Currently, the industry is promoting the construction of prefabricated steel buildings, and the use of stainless steel wall panels for hospitals, office buildings, and large apartment houses. The trend toward panel construction has excited the imagination of many both in and out of the industry.

Changing the Product Mix

Sudden deterioration of the market for a company's products can frequently be a mortal blow from which recovery becomes impossible unless the management is resilient and quickly initiates plans to cope with the problem. A company which was dealt such a blow and has taken steps to overcome the almost disastrous effects of it is Celanese Corporation of America.

As a producer of chemical fibers—acetate and rayon—it originally tied its growth to the textile industry. The inability of its products to hold their own against the onslaught of competition from the "miracle" fibers —nylon, orlon, dacron, and others—led the company to redirect its chemical know-how along other avenues: increasing emphasis on chemical products and plastics, and the development of improved textile products which would compete more successfully with the new "man-made" fabrics. To implement this program, the company instituted a long-range capital plan involving over $100 million in new facilities. It was necessary to include a textile product improvement program to keep the textile operation from dragging down earnings during the slow transition toward chemicals and plastics; in addition to which the company wished to remain an important factor in the field of synthetic fabrics—a field in which the successful development and promotion of new products can be extremely profitable. The company's own success with acetate, and du Pont's fabulous success with nylon (the most commercially important chemical discovery of the last 25 years) is proof of this.

---

[4] The percentage fluctuates, of course. In 1955 it was 23.1 per cent, and in 1956 it fell to 17.8 per cent.

Today, while textiles are still the biggest part of the business, the company seems well on its way to achieving its goal of building an enterprise which would, by 1960, offer a product line comprising about one third each of chemicals, plastics, and textiles. Successful achievement of this goal would improve the profitability of the enterprise, lessen its dependence on an oft-depressed textile industry (advantage of diversification), and raise its standing in the capital markets so that future financing will be less costly.

> The most significant factor in our improved performance results from changes which have been made in facilities, methods and marketing philosophy over the past few years to reorient the Company to meet current and future challenges. The effects of these changes are now beginning to develop.[5]

It is usually a difficult matter, in the planning of capital expenditures, to relate these outlays to a given volume of sales, and for such problems experience is undoubtedly the best teacher. In a growth industry, the problem is somewhat easier to deal with. Thus, plant expenditures for chemical products have, in postwar years, been planned with the expectation of generating a dollar of sales for each dollar of capital outlay; plastics have been expected to bring in about $1.25 per dollar of capital expended. Recently, however, increased competition in the industry has made it more difficult, on the average, to maintain such a favorable relationship between capital outlays and new sales volume.

A Perplexing Problem for a Regulated Industry

The airline industry illustrates still another type of problem in capital planning. In a sense, expenditures for new equipment are "imposed" upon the industry by the rapid rate of technological advances and the psychology of the traveling public. So conscious is the public of speed of air travel that any airline which can offer as much as a five minute differential in flight schedules will take the business away from competing lines. The result is that when an airline orders a new and faster airplane, all the other airlines must acquire the same, or similar, equipment, for existing equipment is made virtually obsolete. Furthermore, the new airplane is typically a substantially more expensive piece of equipment than that which it supersedes. These costs have climbed at a fantastic rate—for example, the Boeing 707 jet, which cruises at 550 miles per hour and carries 100 or more passengers, is available at a cost of $5.5 million. An order of thirty such airplanes (placed by American Airlines) involves a capital outlay that can hardly be treated in an *ad hoc* manner.

Further complicating the financing headaches of the airline companies is the fact that where, in the past, they were able to sell obsolete equipment in the second-hand market at prices which actually resulted in capital gains, the market for the newer equipment is much more limited. Hence, a too-early obsolescence of the new jets by still faster airliners is a costly prospect which currently confronts the industry, and must somehow be planned for.

---

[5] *Celanese Corporation of America*, Annual Report of 1956, p. 3.

Concluding Comments

Forward planning of capital expenditures does not ordinarily reach out as far into the future as does the projection of the product demand on which these expenditures are based. This is partly, at least, a recognition of the time lag that exists between the outlay of capital funds and the sales that the new facilities must generate. Furthermore, the uncertainty factor enters in to cause capital planners to be much less detailed with their projections into the more distant future. Thus, the first quarter's estimates might show monthly detail, followed by only quarterly figures for the remainder of the first year. Then, for the second and third years (and subsequent ones), only annual estimates will be shown.

By and large, capital expenditure plans are kept as flexible as possible, so that any necessary changes can be quickly instituted. However, a "point of no return" is eventually reached—firms dislike to abandon a project once construction has begun, unless there is evidence that the changed picture makes it advisable to take an immediate shorter loss rather than continue with the project and risk a larger one. Thus, in 1957 a number of companies announced the decision to postpone projects which it was felt would be advisedly delayed. In this group were such giants as General Electric, Aluminium, Ltd., and General Motors. On the other hand, Aluminum Company of America, feeling it had gone too far along on a new aluminum facility then in progress, resorted instead to a "stretch-out."

Aside from the cost involved in abandoning a project, the reluctance to do so seems to reflect (appropriately) a recognition that plant facilities are long lived and that temporary adverse movements need not make such facilities unattractive. Thus, where short-run considerations may frequently overly influence managerial decisions before projects have actually been instituted, the underlying long-term factors coupled with the prospect of abandonment costs tend to hold sway once the project has been started.

MAINSPRINGS AND PROBLEM AREAS OF CAPITAL MANAGEMENT

The science of capital management involves several broad and complex phases. One of these concerns the administrative and organizational aspects discussed briefly in the previous section; it involves the establishment of effective procedures for encouraging the creation and facilitating the processing of project proposals in a manner that will make possible the selection of the best ones. As complex as this may seem, it has been, as already summarized, quite adequately handled by many of the better-managed corporations. A second phase deals with the development of a set of principles that will provide management with a guide for selecting from among the available alternatives those projects that would be most suitable and those that would not. Unfortunately, success in this direction has been lacking until relatively recent years, for it is here that the theoretical and analytical aspects of capital management exist, and where the translation of difficult concepts into operational rules are sorely needed. Capital budgeting theory must provide a set of tools which the business decision maker can learn to understand and use in planning his capital budget. Accordingly, it is in this direction that we will concentrate our attention throughout the remainder of this and the following chapter. We may begin by first sketching the background of the science of capital planning and then the practical problem areas to which it gives rise in its application to business management.

Capital Planning Rooted in Capital Theory

The fountainhead of capital budgeting principles is in that branch of economics known as "capital theory"—a body of thought developed over the last eighty years in the writings of such celebrated economists as Boehm-Bawerk, Wicksell, Fisher, and Knight in the earlier period, and more recently by Keynes, Hicks, Boulding, Samuelson, Hayek, and the Lutzes.[6] Capital theory has long been one of the most difficult areas of economics and, for many years, remained outside the main stream of economic thought which traditionally has been concerned with the theory of production and costs in the individual firm. An integration of capital theory with production and cost theory was greatly needed, therefore, as well as an adaptation of the theory to the field of capital management in a manner such that the theoretical principles could be applied to the solution of some of the difficult problems associated with the decisions on the part of businessmen to spend (invest) capital.

Interestingly enough, both of these tasks were accomplished independently and simultaneously, and as recently as 1951. In that year, two important books in the field were published: Friedrich and Vera Lutz, *The Theory of Investment of the Firm*, and Joel Dean, *Capital Budgeting*. The former represents the most complete and modern statement of capital theory currently available. But while it provides the principles needed from which the working tools of capital budgeting can be fashioned, it is probably too theoretical and mathematical in its presentation for most businessmen to comprehend, or for the typical graduate student even to attempt. The latter, on the other hand, is a more readable though less sophisticated exposition, and seeks to shape the existing body of theory into a workable and systematic form that could be presented to the businessman as a point of departure for coping with investment problems.

---

[6] There are many others as well. See also the bibliographical note at the end of Chapter 11.

Thus, while both of the above books represent significant achievements in what is perhaps the most vital area of business decision making, from the point of view of the business administrator troublesome gaps remain. These failings only serve to emphasize the difficulty of translating some of the theory into practice (rather than constituting any oversight or inadequacies on the part of either Dean or the Lutzes) and point up the need for continued research and improvement. In the course of completing this subject we will indicate the areas in which refinement is still needed, and will attempt to make specific suggestions in some instances where suggestions seem appropriate.

Problem Areas

We might begin by asking: What is involved in a request for capital funds to finance a given project? Manifestly, each such request must be supported by data based on a forecast of anticipated costs and revenues over the productive life of the project. The cost data will be based typically on engineering estimates; the revenue data will require a market analysis. If a new product is involved, the market analysis can be a heroic undertaking. In any case, the forecasts will have to be made—and the reader of the earlier chapters of this book has been adequately impressed with the dangers and difficulties involved. We are thus immediately confronted with the first of six problem areas.

1. *Making the Forecast.* The cost and revenue data themselves do not provide a satisfactory enough basis on which the decision maker can act. The correct measuring stick must be employed if the firm is to maximize profits. That the "experts" themselves have not been in accord on what the best measure is, has already been pointed out in Chapter 4 on profit management; and that corporate managements employ crude rules of thumb which often lead to uneconomic decisions will be brought out subsequently.

2. *Determining the Measuring Stick.* The given project must compete for the firm's limited resources with other project proposals. It is, therefore, necessary that management be supplied with some standard for determining which projects to accept and which to reject. That a given project will prove to be profitable may be a necessary, but not a sufficient, condition for accepting it. All of the project proposals will presumably add to or maintain profits, but the scarce capital funds must be rationed, and rationing implies selectivity. The third problem area, therefore, is to determine the acceptance criterion.

3. *Establishing the Acceptance Criterion.* The capital funds may be supplied from a variety of sources within and outside of the firm. These together comprise the firm's potential availability of, or access to, capital. The entire array of project proposals, on the other hand, submitted to top management, constitutes the demand for capital funds. These demands for capital might, for the large part, be acceptable accord-

ing to the acceptance criterion, but not in terms of capital availability and other factors that call for top management's consideration, e.g., dividend policy and capital structure. Hence, some projects will be temporarily shelved, to be instituted, if still desired, at a later time. This leads to the planning and budgeting of capital resources.

4. *Planning the Capital Budget, Short-Term and/or Long-Term.* With a given budget size for the ensuing fiscal year and even for the next few years, the alternative sources of capital raise the question of financing methods. The many instruments of finance themselves reflect the tastes and needs of the many seekers of capital. The various types of equity and debt financing that might be resorted to thus constitute another problem area.

5. *Determining the Methods of Finance.* The difficulties encountered in each of the above problem areas are compounded by the uncertainty factor. So all-pervasive is this, and so important, that failure to give it adequate recognition will usually lead to faulty and costly decisions. Thus, despite the fact that uncertainty is a part of each of the problem areas listed above, it is important enough to warrant separate listing, for it is in this area that the greatest need for research and the most room for improvement remain.

6. *Coping with Uncertainty.* As will be made clear shortly, the six "trouble spots" are not single problems but involve, in each case, a problem complex—a set of interacting forces the effects of which cannot be readily traced. The first problem complex, that of making the forecast, has already been discussed in detail in earlier chapters and will, therefore, receive no further treatment here. The uncertainty problem, on the other hand, which was treated in the first chapter, will come in for continued discussion, but not as a separate topic. Instead, it will be woven in with the discussion of how to cope with the other problems encountered in capital planning. As will be brought out subsequently, we shall find that the establishment of an acceptance criterion is actually the most difficult of the problems encountered in the field of capital budgeting (the forecasting problem is, of course, a monumental one, but it is not unique to capital budgeting), and constitutes almost virgin territory in this field. Much work remains to be done; and we have found, in our own explorations, that it was necessary to tie the discussion in with a consideration of financing methods. Hence, the latter topic will not be discussed as an independent problem complex. Finally, the fourth problem area, that of planning the capital budget, has in various respects already been discussed, and will come in for further treatment in other sections which follow.

DETERMINING THE MEASURING STICK

Corporate managements have naturally been interested in using a tool, in their capital planning problems, which could be readily understood, deftly employed, and which would produce results that were not too different from those yielded by more sophisticated and time-consuming techniques. Consequently, a number of "rule-of-thumb" methods have been, and remain, in vogue in many business enterprises; in other cases the cruder methods have been replaced by more sophisticated techniques. We will discuss, in this section, the more widely used methods and will indicate what seem to be pragmatic approaches to the problem of determining a suitable measuring stick.

Urgency or Postponability

This is among the crudest of the methods that management might employ. Because it is more qualitative than quantitative there exists the tendency to put off, indefinitely, projects which might actually be highly profitable, simply because it is easy to put them off or because the officer who has proposed a less profitable project is better able to convince the executive committee of its "urgency." It may frequently be true that a very urgently needed project will offer a high rate of return so that it is inconsequential whether the project is accepted because of its urgency or its profitability—the result is the same. But coincidence of this sort is not very reliable.

The Pitfall of "Double Counting." The urgency criterion can also lead to pitfalls which sophisticated methods are likely to avoid. For example, there are hardly more urgently needed investments than those which involve the replacement of a worn-out part of a profitable facility, without which part the facility could not function: an engine for a commercial airliner, a generator in a utility plant, a blast furnace in a steel mill, a kiln in a cement plant, and so on. Failure to replace any integral part of the respective facilities when they break down (or to replace any of the smaller components which are vital to the functioning of these integral parts) means that the revenues and profits from the entire facility will be foregone. On the face of it, the conclusion seems obvious that urgency and high profitability[7] are coincident in this case. That this is an erroneous conclusion might best be made clear with our example of the commercial airliner.

Soon after the worn-out engine is replaced with a new one, a second of the four engines breaks down, and again the profits of this facility can be rescued at relatively small cost. The process will be repeated when the other engines must be replaced, when the electrical system must be repaired, and when the interior of the plane must be refurbished to provide passengers with comforts on a par with competing airlines. Each repair and replacement results in a rescuing of profits, but it is always the same profits that are rescued. The fallacy of this approach lies in the failure to view the profitability of the facility over its productive life and to estimate *all* the costs that will be necessary to produce the future revenues. Only when all future costs and revenues are estimated can an economic decision be made as to whether the particular profits are worth rescuing or whether the investment funds required should instead be channeled into other opportunities, i.e., a new commercial airliner.

---

[7] The profitability of the additional investment is apparently very high, for it involves the replacement of only a part of the total facility; and when we compare this replacement cost with savings (profits) made possible by putting the facility back into operation, we get an extremely high rate of return on the added cost.

In a rather unprecise manner, the passenger car owner will frequently go through this very procedure. When confronted with the need to replace the tires on his car he will mentally compare the future revenues (in the form of services) that the car is still capable of producing, with the estimated outlays (usually only major ones) that can be expected —probable relining of the brakes, new battery, new clutch, perhaps an engine overhaul—and he will then either buy the new tires or trade in his car for a newer model.

Payout, Payoff, or Payback Period

Businessmen have frequently employed a short-cut method of allocating capital funds by estimating the length of time required for the cash earnings on a given investment to return the original cost to the owner. This measure is referred to in various parts of the literature as either the payout, payoff, or payback period. It is used both as a before- and after-tax measure (the latter being the more significant), and its expression in terms of cash earnings is a recognition of the fact that depreciation and depletion charges should be included in the earnings figures, i.e., earnings are measured before depreciation and depletion.

By way of simplifying the employment of this tool, it is typical to estimate a uniform flow of annual earnings over the life of the project. Hence, if the original investment is represented by I and the uniform (average) annual cash flow is represented by E, the payout period, P, is expressed as:

$$P = I/E \,. \qquad (1)$$

Under the conditions stated above it is clear that, given the life of the project, profitability will vary inversely with the payout period, and that, given the payout period, profitability will vary directly with the life of the project. It is, therefore, easy enough to understand the insistence of management on short-payout investments. However, the tool is often too blunt to be used in selecting among alternative projects which differ as to cost, payout, and productive life. In such cases, a more precise instrument is needed. It is worth pointing out, at the risk of being obvious, that a short payout is not necessarily coincident with high profitability. Thus, if the productive life of the project is even shorter than the payout period, the return on the investment will be negative; and if the project life and payout period are equal, the investment return is zero. In either case, an economic loss is incurred.

Effect of Corporate Income Tax on Payout. The relationship between the pre-tax and after-tax[8] payout depends both on the tax rate and the productive life of the project. Representing depreciation by D and assuming a 50 per cent tax rate, the after-tax payout becomes:

$$P' = \frac{2I}{E + D}. \tag{2}$$

As a limiting value, for projects of perpetual life (or, for practical purposes, extremely long-lived projects such as hydroelectric plants and dams), the depreciation charge approaches zero and the after-tax payout approaches twice the value of the pre-tax payout. Hence, with a 50 per cent tax rate, the after-tax payout will, for all depreciable investments, lie somewhere between the pre-tax payout at the lower limit and twice that value at the upper limit.

Blunt as the tool may be for many purposes, the payout period concept is extremely important if only because it is used so widely in business. However, this is not its only claim to fame, so, while we shall at this point turn to a consideration of other measuring sticks, we will have more to say about the payout tool in the discussion of the other methods below.

Capital Recovery Period

Where the payout period is concerned merely with the return of the original investment, the capital recovery period takes account of the time value of money (interest) as well. Assuming again a uniform periodic repayment, the capital recovery period expresses the length of time in which a given investment will be paid off at some assumed rate of return. Thus, if:

U = the uniform periodic repayment,
r = the rate of return (interest) per period,
I = the original investment, and
n = the number of periods over which the uniform repayment U must be made in order to recover the investment I at the desired rate of return r,

then:

$$n = \frac{\log\left(\dfrac{U}{U - rI}\right)}{\log\ (1 + r)}. \tag{3}$$

Equation (3) may be referred to as the uniform capital recovery formula, and its derivation is given in the section below. Since it is a rather cumbersome formula to work with, its use has been facilitated by the construction of financial tables so that it is a relatively simple matter to turn to the correct page of the tables and read the desired answer.[9]

---

[8] Discussion of an after-tax payout seems meaningful only for projects which are actually profitable, i.e., the productive life is greater than the pre-tax payout period.

Capital Recovery versus Payout. If the interest rate were zero, the capital recovery period would be the same as the payout period. Since the time value of money is obviously greater than zero, the capital recovery period for a given investment project will exceed the payout period if the dollar amounts of the periodic returns, U and E respectively,[10] are equal.[11] Conversely, for any given cash flow estimated into the future, the required capital recovery period may be made as short as desired by specifying a low enough rate of return on a given project.

Defining the Rate of Return. Employed skilfully, the capital recovery period can be a useful tool for certain types of investment decisions. Where the investment will have little or no effect on the revenue stream, but is aimed at cost reduction or replacement of obsolete equipment, this tool provides the means for making the correct economic decision in choosing from among two or more alternatives. At this point in the presentation, however, the use of this measuring stick raises one troublesome question which we will have to postpone answering: How do we determine the rate of return, r? For the moment we will say that r represents the minimum rate of return that the entrepreneur is willing to accept.

Annual Cost

As presented above, the capital recovery formula does not find wide application. The stress on short-payoff and short-capital-recovery periods is a reflection of the businessman's desire to minimize risk by not having to expose his investment to what he feels is too long a wait to get back his outlay.

---

[9] In his article, "How to Figure Equipment Replacement," *Harvard Business Review* (September, 1955), p. 81, P. A. Scheuble, Jr. illustrates the use of a device called a *nomogram* which relates the rate of return, capital recovery period, original cost, and annual savings (cash earnings). This device provides a quick, approximate method for determining any one of the above factors when all the others are known. However, it affords little or no advantage to one who has gained facility in the use of capital recovery tables.

[10] The reason for using different symbols in distinguishing between payout period (P) and capital recovery period (n), and between uniform cash earnings flow (E) and uniform periodic repayment (U), is that the respective concepts are really so very different, even though, superficially, they seem to be quite similar. The difference is more than merely that of a zero interest rate versus a rate greater than zero. The payout concept abstracts from the interest rate altogether; the capital recovery concept makes explicit use of it. Thus, payoff is dependent only on the (uniform) flow of cash earnings relative to the original investment; capital recovery depends as well on the required rate of return.

[11] Because U is made up of both principal and interest. Hence, if $U = E$, a greater period of time is required over which U must flow as contrasted with E in order for the given project to pay off.

However, a wise economic choice among investment alternatives can more effectively be made when the decision is based on what is called "annual cost" rather than on capital recovery periods. This involves the use of exactly the same formula, but instead of solving for n we solve for U. Again the solution is arrived at by employing appropriate financial tables, although more pioneering spirits might be tempted to solve mathematically the expression:

$$U = \frac{Ir(1 + r)^n}{(1 + r)^{n-1}}. \tag{4}$$

To repeat, this is exactly the same formula as the one stated in equation (3), except that the solution is for U. Stated nonmathematically, the formula says that if an investment equal to I were made today (or at any time in the future) having an estimated life of n years, and if the minimum required rate of return were r, then the project would have to produce over its lifetime a cash earnings flow which would be equivalent[12] to a uniform annual flow of U. Stated still another way, U represents the uniform annual payment necessary to repay the investment, I, at the minimum required rate of return, r, over the life of the project, n.

Application to Real Estate Mortgage Loans. While this concept as stated may seem unfamiliar to many, it is actually among the most commonly lived-with concepts of business. The typical homeowner's mortgage payments are determined by this very formula as expressed above. The bank (or other lender) takes a mortgage on the home in question (to the lender this is the investment, I) and stipulates the rate of return, r, (interest) at which I is to be repaid, and the number of years, n, over which the uniform payments, U, will flow. In this particular form of investment, the uniform periodic payments are usually made monthly rather than annually, but this does not alter the basic concept.

If the rate employed in the formula is truly the minimum return that the investor would be willing to accept, then U represents a uniform periodic cash flow which just compensates him for his present investment, I. In other words, the investor is indifferent as between a present sum of money, I, and a uniform *annual cash* flow, U, for n years. Hence, to the investor U represents the annual cost of the investment.[13] Should the investment involve periodic disbursements for maintenance and repair, power, insurance, property taxes, etc., these disbursements must be estimated and added to U as a measure of the annual cost of the given investment, for U represents only the payment which goes toward the recovery of the original outlay (at interest).

---

[12] Note that the actual cash flow would almost certainly be an irregular one, but this stream is, nevertheless, equatable to a stream comprising a uniform, periodic flow of U payments.

[13] Here again it is seen how the concept of opportunity cost is very much a part of business life.

Application to Installation of Equipment. As already stated, the capital recovery formula is best suited to helping the investor select from among alternatives which are expected to affect only the cost of producing a given product or service while having little or no effect on the revenue stream. Such decisions are typically involved when the question of equipment replacement is raised, or when the choice must be made between installing one type of machine as against another. Furthermore, the use of the formula is necessary only where there is no obvious advantage of one alternative over the others. Thus, if machine A requires a smaller outlay than machine B, has at least as long a life, costs no more to operate and maintain, and does just as good a job as the latter, obviously the wise choice is to use machine A rather than machine B.

But suppose machine A involves an initial outlay of $10,000 and will require additional annual monetary outlays for operation, maintenance, etc., of $2,500. On the other hand, machine B costs $13,000 but is much more economical to operate, requiring disbursements of only $1,800 per year. Assuming that the machines both have an economic life of eight years, that the minimum required rate of return on the investment is 12 per cent, and that the machines will have no salvage value at the end of their economic lives, which of the two machines would be the best economic choice? The answer is obtained by solving equation (4), or, much more simply, by reading the correct figure from the appropriate table. Thus, the investment of $10,000 in machine A requires an annual capital recovery payment (U) of $2,013, while the investment in machine B requires a payment of $2,617. When we add to these capital costs the annual disbursements for operation and maintenance we find that machine B involves a smaller annual cost—$4,417 as against $4,513 for machine A. Hence, management should install machine B.

Derivation of Capital Recovery Formula. While knowledge of the derivation of the formula is in no way necessary for its effective employment, since the usual and really sensible method is to read the answers in tables designed for the purpose, an understanding of how the formula comes into being will provide the key as to why it works. The reader with no more than a vague recollection of high school algebra should be able to follow the logic, though it is not necessarily the sort of thing that one need try to commit to memory.

Since U is a payment that is to be made available to the investor at the end of each year over the life of the investment, the first payment received at the end of the first year can be invested for the remainder of the $(n - 1)$ years at a rate of interest, r. From the discussion of compound interest in Chapter 4, it will be recalled that under these conditions the investment will grow to a value of $U(1 + r)^{n-1}$ by the end of the project's life. Similarly, the payment, U, received at the end of the second year can be invested for $(n - 2)$ years and will grow to a value of $U(1 + r)^{n-2}$. The final payment, U, made at the end of the nth year will have, of course, a value of U at that time, while the payment received one year before the end of the project's life will have earned interest for that year and grown to $U(1 + r)$. Thus, the entire series of payments made over the life of the project will have a total value equal to the sum of the values to which the individual payments will have grown, and will be equal to:

$$U(1 + r)^{n-1} + U(1 + r)^{n-2} + \ldots + U(1 + r) + U \,. \qquad \text{(a)}$$

However, instead of making the investment in the first place and accepting a periodic flow of payments amounting to U per period, the investor could have put his money, I, at compound interest for n years so that at the end of that time his investment would have grown to $I(1 + r)^n$. Hence, the uniform periodic payment, U, must be just large enough so that if it were invested periodically upon receipt, it would make the investor just as well off as if he had invested his original sum, I, at compound interest. Thus:

$$U(1 + r)^{n-1} + U(1 + r)^{n-2} + \ldots + U(1 + r) + U = I(1 + r)^n. \qquad \text{(b)}$$

Now with some "conjuring" tricks we get the desired result. First we multiply both sides of equation (b) by $(1 + r)$, so that it becomes:

$$U(1 + r)^n + U(1 + r)^{n-1} + \ldots + U(1 + r)^2 + U(1 + r) = I(1 + r)^{n+1}. \qquad \text{(c)}$$

Then we subtract equation (b) from equation (c). The reason for this maneuver is that many of the terms on the left-hand side of equation (b) appear also in equation (c), so that the subtraction process serves to eliminate all the terms common to both equations. The result is the following simplified expression:

$$U(1 + r)^n - U = I(1 + r)^{n+1} - I(1 + r)^n. \qquad \text{(d)}$$

We need now merely reduce this equation by a process of factoring and regrouping of terms, then solving for U. When this has been effected, the resulting equation is:

$$U = \frac{Ir(1 + r)^n}{(1 + r)^{n-1}}\,. \qquad (4)$$

Equation (4) is the basic capital recovery formula, and by restating it as a solution for n (the capital recovery period) we get equation (3), as presented earlier in this chapter.

Rate of Return

The common (lay) meaning of rate of return on investment is simply the ratio of annual receipts to original cost. This definition is only approximately correct, and becomes precisely true for a permanent, nondepreciating, nonappreciating asset, producing a periodically uniform income stream. To the extent that the liquidating value or resale value of the asset is, at the time of liquidation or sale, greater or less than the original outlay, the true rate of return will be greater or less than the rate as defined above. Since at the time the investment is made one rarely is able to predict its precise liquidating or resale value, it follows that in the typical case the *true* rate of return on an investment to a given owner cannot be known until the ownership has terminated. Hence, prior to actual termination of ownership, it is necessary to accept the best estimated rate of return as a reasonable measure of the investment's profitability.

Generalized Definition of Rate of Return. The precise definition of rate of return on any investment is: that rate which equates the present value of the cash receipts expected to flow from the investment over its lifetime with the present value of all expenditures relating to the investment. To minimize unnecessarily complicating aspects without doing violence either to the concept or the conclusions, it is usual to assume that the investment involves only an initial outlay of funds. Where additional outlays are expected to be required in the future, however, these are simply discounted down to the present and added to the original outlay to determine the total present value of outlays. Thus, the total cost of the investment may be expressed as:

$$I = C + \frac{O_1}{1 + r} + \frac{O_2}{(1 + r)^2} + \ldots + \frac{O_n}{(1 + r)^n}, \qquad (5a)$$

where C is the initial cost outlay, O_1, O_2, . . . O_n are a series of future outlays, and I is the sum of all outlays properly discounted to represent the present value of the investment.

The project will also produce a flow of cash earnings over its life which must be similarly discounted down to the present.
Thus:

$$I = \frac{R_1}{(1 + r)} + \frac{R_2}{(1 + r)^2} + \ldots + \frac{R_n}{(1 + r)^n} + \frac{S}{(1 + r)^n}, \qquad (5b)$$

where R_1, R_2, . . . R_n are a series of cash flows received at the end of each of the respective periods over the life of the investment, S is the liquidating or salvage value at the end of n periods, and I is the present value of this stream of future receipts.

Equation (5a) expresses the present value of the investment in terms of the outlays connected with it; equation (5b) expresses the investment's value in terms of the cash revenues that will flow from it. If we select that rate of discount, r, in such a way that I will be the same in both equations, as we have tacitly done above, then r is defined as the true rate of return on the investment.

Simplified Version of Rate of Return. For analytical purposes, equations (5a) and (5b) are too cumbersome to work with, so the analyst usually makes certain simplifying assumptions which, as stated above, do no violence either to the rate of return concept or to the conclusions reached. The typical assumptions are:

1. The original cost is the total cost (no future outlays will be required), so that equation (5a) reduces to $I = C$.
2. The cash earnings flow is uniform, so that we simply represent each year's cash flow by a uniform receipt, U.
3. The salvage value is small relative to the original outlay and the present worth of the salvage value, being still less, may be ignored, so that the term $\frac{S}{(1+r)^n}$ in equation (5b) drops out.

From the last two of the above assumptions, equation (5b) simplifies to:

$$I = \frac{U}{1+r} + \frac{U}{(1+r)^2} + \cdots + \frac{U}{(1+r)^n} \tag{5c}$$

and is easily recognized as a geometric progression in which each term differs from the preceding one by a factor of $\frac{1}{(1+r)}$. This enables us to apply the formula for the sum of a geometric progression to equation (5c), giving us:

$$I = \frac{\frac{U}{1+r}\left[1 - \left(\frac{1}{1+r}\right)^n\right]}{1 - \frac{1}{1+r}} \tag{5d}$$

which simplifies down to:

$$I = \frac{U}{r} - \frac{U}{r}\left(\frac{1}{1+r}\right)^n. \tag{5e}$$

Equation (5e) is to be interpreted as saying that for an investment, I, from which it is estimated that there will be a uniform annual cash flow of U, the true rate of return on this investment is that value of r which will equate the total cash flow with the investment outlay.

For any specific problem to which the foregoing simplifying assumptions do not apply, equation (5e) should not be used—obviously. In such cases it will be necessary to find that value of r which equates equation (5a) with equation (5b). (Note: setting the right-hand side of equation (5a) equal to the right-hand side of equation (5b) constitutes a mathematical expression of the precise meaning of rate of return.) But for analytical purposes, equation (5e) is far the more desirable form, and will be applicable to a wide range of practical problems as well.

The Rate of Return and Allied Concepts. The rate of return on investment, technically defined by the set of equations (5a) and (5b), is a well-conceived theoretical concept that goes by various "aliases" in the economic literature. Among the more important of these are "marginal efficiency of capital," a phrase made famous by John Maynard Keynes,[14] and the "internal rate of return," used by Kenneth E. Boulding[15] and others. But while it is a concept well founded in theory, it presents many practical problems to anyone desiring to apply it to actual cases, for it involves a great amount of trial and error work if one were to insist on solving for the precise rate of return from any of the applicable equations discussed above. Such precision is actually unnecessary and would, in fact, be misleading, for we must realize that the solution for r depends on the value assigned to (estimated for) U. Hence, the rate of return can itself only be an estimate, so that insistence on a precise solution for r would be unrealistic. It is worth repeating at this point that the rate of return on an investment can never be stated with precision until the ownership of the investment in question has been terminated.

Approximations to the Rate of Return. Because precision is often impracticable and, as indicated above, unrealistic, shorthand approximations are usually employed.

The Payout Reciprocal. Let us first rewrite equation (5c) by solving for r. We then get:

$$r = \frac{U}{I} - \frac{U}{I}\left(\frac{1}{1+r}\right)^n . \qquad (6)$$

Since U is the uniform annual cash flow, and I is the investment outlay, it is clear that $\frac{U}{I}$ is the reciprocal of the payout period. It follows, then, that the rate of return is the difference between the reciprocal of the payout period and some quantity equal to the product of the payout reciprocal and $\frac{1}{(1+r)^n}$. And it is immediately obvious that for large values of n (long-lived projects) this second quantity will be small and, therefore, the rate of return will be approximated rather closely by the value of the payout reciprocal.

For a project whose life is permanent (or practically so, as a hydroelectric dam and other very-long-lived investments) n is, of course, infinitely large and the value of $\frac{U}{I}\left(\frac{1}{1+r}\right)^n$ becomes zero, so that the return on such investments is exactly equal to the reciprocal of the payout period.

It thus seems rather interesting that a rule-of-thumb method employed by businessmen for a great many years should actually prove to have been a reasonably good approximation of the theoretically correct measure. Furthermore, this method has a rather wide applicability to business investment problems, for the rate at which the factor $\frac{1}{(1+r)^n}$ approaches zero, as n becomes large, increases rapidly as r itself increases. Since few businessmen would be willing to consider a project offering anything less than 20 per cent (before taxes), the payout period (and its reciprocal) becomes, for practical purposes, a very handy tool.

---

[14] *General Theory of Employment, Interest, and Money.*

[15] *Economic Analysis,* 3d ed., chap. 39.

Employment of the tool requires recognition of the fact that the *payout reciprocal is always a maximum estimate of the true rate of return*, for, as pointed out above, the solution for r from equation (6) is obtained by subtracting some quantity from the payout reciprocal. Hence, in using the payout reciprocal as an estimate of the rate of return we are ignoring the quantity which must be subtracted. This quantity may properly be ignored, in theory, only when n is infinitely large; but, in practice, the payout reciprocal would appear to be a very satisfactory estimate of r in all cases where the project life is "substantially" greater than the payout period. We will see shortly that where n is less than twice the payout period[16] the error of estimate is too large to be ignored. In such circumstances, another approach would be preferred. Happily, the great majority of projects likely to be given serious consideration in the typical business enterprise are those for which the payout period is very short relative to the project life, so that the payout reciprocal is, practically speaking, an admirable tool for quick estimation of a project's rate of return.

The Outlay-Revenue Ratio. A more accurate, but much more laborious, method of estimating r is an approach which employs the trial and error technique. But for the fact that a precise solution for r is unnecessary and would, in fact, convey a false idea of the degree of advancement in the science of prediction, the trial and error method would be a distastefully lengthy procedure. However, a willingness to be content with reasonable approximations makes possible the use of various short cuts. Nevertheless, the method to be discussed should be adopted only when the payout reciprocal is deemed unsatisfactory: (1) where the estimated project life is less than twice the payout period, or (2) where, for reasons known to the investigator, a more precise estimate is desired than can be made from the payout reciprocal.

The principle which underlies the particular trial and error method explained here stems from the realization that if the true rate of return is that rate which equates the present value of outlays with the present value of cash earnings over the life of the project, then the ratio of these flows, when discounted at the true rate of return, is equal to unity. To determine this rate, it becomes simply a matter of setting up a series of columns, as has been done in Table 10–1, on work sheets that have been specifically prepared for the purpose.

---

[16] *Cf.* M. J. Gordon, "The Payoff Period and the Rate of Profit," *The Journal of Business* (October, 1955), p. 253. Professor Gordon suggests the suitability of an n as low as 1.5 times the payout period, but this, as will be shown shortly, can lead to great divergencies from the true rate of return.

In the first column are the years over which the cash flows of the second column are estimated. Outflows (negative items) of cash constitute the actual investments that make up the project and, in the current example, these are expected to be made in one single expenditure in the present. The revenues are assumed to begin one year later and to flow into the enterprise in a uniform stream at each subsequent year end. Projects of varying lives have been treated in the table—10 years, 12 years, 13 years, and 15 years—and the solution for their rates of return is derived from the data on lines A, B, C, and D respectively. That is, on each of these lines are given the outlay-revenue ratios, O/R, obtained by first discounting the outlay and revenue quantities at the rates indicated in the headings of the various present-value columns.

The basic cash flow data are given in the second column, headed "Cash Flows." The O/R ratios for this column have been computed, for each of the project lives (10, 12, 13, and 15 years respectively), by dividing the total outlay of $10,000 by the full-life stream of respective revenues. Thus, for the project with a 10-year life, the O/R ratio shown in the basic "Cash Flows" column is 0.67, determined by dividing the total outlay of $10,000 by the total 10-year stream of revenues of $15,000. In effect, we have discounted all data in this column at a rate of zero per cent, so that the estimated future dollar flows have the same value in the present as they will have in the future. In the succeeding columns, since the discounting rates are all greater than zero, the present values of the future dollar flows will all be smaller than the corresponding quantities shown in the basic "Cash Flows" column. These present values (abbreviated in the column headings as PV) are determined by multiplying each of the cash flow items by the appropriate discount factor (abbreviated as DF). Thus, at a discount of 8 per cent, a dollar due one year from today has a present value of $0.926. Using 0.926 as the discount factor, we simply multiply it by $1,500 to find that the latter sum, due one year from today, has a present value of $1,389 when discounted at 8 per cent.

Since published tables of discount or present worth factors are available, it is possible to effect a time saving with this operation by having work sheets prepared on which the discount factors for various rates of interest have been placed. It then becomes a matter for some junior employee to convert the cash flow data into present values in each of the PV columns, with the aid of a calculating machine. The O/R ratios are then determined by dividing the present value of outlays by the respective present value of revenues. In the illustration, the single outlay item of $10,000 is multiplied in each case by a discount factor (DF) of 1.000 because the outlay is assumed to have been made in the present.

After computing the various O/R ratios it is a relatively easy matter

TABLE 10-1

Rate of Return

Year Ending	Cash Flows	Present Values of Cash Flows at Various Rates of Discount											
		8%		10%		12%		15%		20%		25%	
		DF*	PV†	DF	PV	DF	PV	DF	PV	DF	PV	DF	PV
0	$10,000	1.000	$10,000	1.000	$10,000	1.000	$10,000	1.000	$10,000	1.000	$10,000	1.000	$10,000
1	1,500	0.926	1,389	0.909	1,363	0.893	1,339	0.870	1,305	0.833	1,249	0.800	1,200
2	1,500	0.847	1,270	0.826	1,239	0.797	1,195	0.756	1,134	0.694	1,041	0.640	960
3	1,500	0.794	1,191	0.751	1,126	0.712	1,068	0.657	985	0.579	868	0.512	768
4	1,500	0.735	1,102	0.683	1,024	0.635	952	0.572	858	0.482	723	0.410	615
5	1,500	0.681	1,021	0.621	931	0.567	850	0.497	745	0.402	603	0.328	492
6	1,500	0.630	945	0.564	846	0.507	760	0.432	648	0.335	502	0.262	393
7	1,500	0.583	874	0.513	769	0.452	678	0.376	564	0.279	418	0.210	315
8	1,500	0.540	810	0.467	700	0.404	606	0.327	490	0.233	349	0.168	252
9	1,500	0.500	750	0.424	636	0.361	541	0.284	426	0.194	291	0.134	201
10	1,500	0.463	694	0.385	577	0.322	483	0.247	370	0.161	241	0.107	160
	$15,000		$10,046		$9,211		$8,472		$7,525		$6,285		$5,356
A O/R	0.67		0.96		1.09		1.18		1.33		1.59		1.87
11	1,500	0.429	643	0.351	526	0.287	430	0.215	322	0.135	202	0.086	129
12	1,500	0.397	595	0.319	478	0.257	385	0.187	280	0.112	168	0.069	103
	$18,000		$11,284		$10,215		$9,287		$8,127		$6,655		$5,588
B. O R	0.56		0.89		0.98		1.08		1.23		1.50		1.79
13	1,500	0.368	552	0.290	435	0.229	343	0.162	243	0.093	139	0.055	82
	$19,500		$11,836		$10,650		$9,630		$8,370		$6,794		$5,670
C O/R	0.51		0.84		0.94		1.04		1.19		1.47		1.76
14	1,500	0.340	510	0.263	394	0.205	307	0.141	211	0.078	117	0.044	66
15	1,500	0.315	472	0.239	358	0.183	274	0.123	184	0.065	97	0.035	52
	$22,500		$12,818		$11,402		$10,211		$8,765		$7,008		$5,788
D O/R	0.44		0.78		0.88		0.98		1.14		1.43		1.73

* Discount factor
† Present values

to determine the rate of return. Since the true rate is that which equates the discounted values of revenues with the discounted values of outlays, we merely look for the O/R ratio which has a value of 1.0. This, we can see, lies between the 8 and 10 per cent columns (for project A, having 10-year life), and even a quick glance reveals the fact that the O/R ratio would be unity at a rate somewhere between 8 and 9 per cent.

For most practical purposes this would be a workable solution, so that it would not be necessary to attempt further refinement. It would even be quite unnecessary to bother to compute all the data entered in the 12 per cent (and over) discount columns, but these columns are shown for reasons to be made clear shortly.

Interpolating for Precision. In the interest of nicety and for aesthetes who cannot be satisfied with anything so unprecise as the statement that the rate of return lies "somewhere between 8 and 9 per cent," a more exact (though not mathematically precise) solution is readily achieved by interpolation. This gives an answer of 8.6 per cent.

Lines B, C, and D indicate the O/R ratios for projects involving the same original outlay, the same annual cash flow, but of different (longer) life respectively. As is to be expected, the rate of return increases with the life of the project since there are additional years over which cash earnings can flow, for the same original outlay. Thus, project B returns approximately 10.4 per cent, C gives 11.2 per cent and D yields a return of about 12.4 per cent.

If the project's life were to be increased indefinitely, providing each year the same cash flow of $1,500, the rate of return would approach 15 per cent. This, as we have seen, is the reciprocal of the payout period $\frac{\$1{,}500}{\$10{,}000}$. Since all of the projects in the illustration involve the same original investment and the same annual cash flow, they all have the same payout period and the same payout reciprocal. The latter, we have seen, is 15 per cent; the former is 6.67 years.

Evaluating the Payout Reciprocal

We can evaluate now the effectiveness of the payout reciprocal as an estimate of a project's rate of return. This might best be done by summarizing our findings in short tabular form, as in Table 10–2.

From Table 10–2 we may conclude that where the project life is

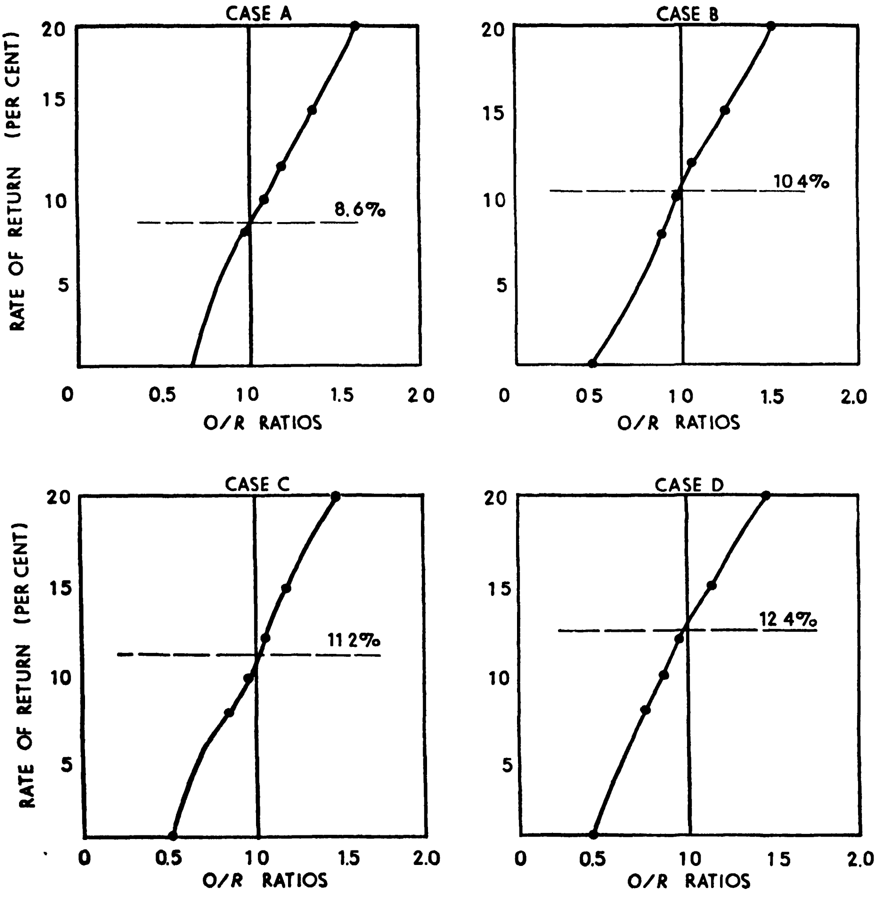
TABLE 10–2

Evaluation Table

Project	Project Life, Years	Life Payout Ratio	Payout Reciprocal, %	Rate of Return, %
A.	10	1.5	15	8.6
B.........	12	1.8	15	10.4
C.	13	1.95	15	11.2
D..	15	2.2	15	12.4

less than twice the payout period, the payout reciprocal can give a very poor estimate of the rate of return, and that certainly for a life-payout ratio of only 1.5, the payout reciprocal is an almost useless estimate of the rate of return. But where the ratio is greater than 2.0, the payout reciprocal provides a good workable estimate, if not always a very close one, of the rate of return. Furthermore, where it is desired to achieve a precision greater than that likely to be afforded by the payout reciprocal, the latter is still a very useful first approximation. Thus, in the problems illustrated here, it is obviously pointless, once the payout reciprocal has been computed, to use discount rates of more than 15 per cent, for we know from our earlier discussion in this chapter that the payout reciprocal represents a maximum estimate of a project's rate of return.[17] Having computed

FIGURE 10–1



---

[17] If the projected revenue stream is such that the cash earnings expand rapidly in the years following the payout period, then the payout reciprocal will not be a maximum estimate. But such situations do not seem too likely, and may well be treated as exceptions which prove the rule.

the payout reciprocal, it is then easy enough to narrow down the range within which the rate of return will lie.

Precision by Graphic Means

As an alternative to interpolating between two O/R values, it is possible to read the rate of return from a graph on which has been plotted the O/R values on the *x* axis, and the various rates of return on the *y* axis. Where the resulting curve (which connects the several points plotted on the graph) intersects the O/R value of 1.00, we simply read off the corresponding rate-of-return value. This procedure is illustrated in Figure 10–1, and nothing more need be said about it other than that it is an unnecessarily time-consuming operation and does nothing to improve on the interpolation technique.[18]

SUMMARY

Of the numerous measuring sticks available, the urgency or postponability method is the least reliable and should be studiously avoided. The payoff period (or its reciprocal), long used by businessmen as a rule-of-thumb method, proves to be a very valuable tool either as an actual estimate of the rate of return where the project life is more than twice the payout, or as a first approximation and guide to determining a closer estimate where the payout reciprocal is itself considered too crude.

For projects that are expected to have little or no effect on the revenue stream, the annual cost method or the capital recovery period are well suited as measuring sticks. These are actually alternative methods of arriving at the same conclusion, since they employ the same (capital recovery) formula. However, because it lends itself to more meaningful interpretation for investment purposes, the annual cost measure is usually preferred to that of the capital recovery period. While we have not taken the trouble to prove it, the investigator would find that, for projects of the type under discussion in this paragraph, the rate of return measure would favor the selection of the same alternatives as those indicated by the annual cost and capital recovery measures if, in the latter applications, the true rate of return were used as the discounting rate. The discussion of what the proper discounting rate should be in using the capital recovery formula has, however, been deferred, to be discussed in the following chapter, and we have agreed, for the time being, to use the "minimum acceptable rate of return," whatever that might be.

Where the actual rate of return must be closely estimated, trial-and-error methods become, unfortunately, necessary. However, the distastefully time-consuming aspects of these methods can be minimized by employing the outlay-revenue approach, together with the payout reciprocal as a first approximation in the estimating process. Where not too great precision seems necessary, a fairly quick, rough estimate can be made by calculating an O/R ratio employing a discount rate of a few per cent under the payout reciprocal. By comparing this O/R ratio with the payout reciprocal and with the O/R ratio obtained from the basic cash flow data, a fairly reasonable estimate can be quickly obtained. If a greater degree of precision is required, the O/R ratios, as they are computed, provide guideposts for narrowing down the range within which the true rate of return must lie. When this range has been narrowed down to a 2 or 3 per cent bracket, interpolation will then provide a very close estimate of the precise rate of return.

---

[18] This technique is employed by R. Reul, "Profitability Index for Investments," *Harvard Business Review* (July, 1957), p. 116.

Other measuring sticks are available which, for want of space, have been neglected. The one all-embracing measuring stick, rate of return, is applicable to all types of projects whether they are expected to affect revenues alone, costs alone, or both. Where some other technique is employed, therefore, it is done only because it provides a short cut for arriving at the same answer to which the rate-of-return measure would point. Hence the emphasis on annual cost and payout reciprocal. Similarly, other methods not discussed here aim at circumventing the difficulties encountered in the rate-of-return method. Thus, the method called "capitalized cost" or "present value" is actually a perfect substitute for rate of return, and employs the same discounting procedure and produces the same answers. The "average investment" method is another approximating formula that can be used as a substitute for payout reciprocal, but it does not seem as useful as the latter. And, finally, we shall mention the famous "MAPI formula," which we have decided to ignore because of lack of space, its limited applicability to replacement investments only, and our belief that it does nothing for the decision maker that the methods discussed here will not do.

BIBLIOGRAPHICAL NOTE

See the end of the next chapter.

QUESTIONS

1. It is said that capital planning proposals flow along a "two-way street." Distinguish between the two types of proposals most likely to flow in each of the two directions.
2. Outline in brief a step-by-step procedure by which a capital expenditure proposal might be initiated and carried through the process of final approval and actual authorization of the expenditure as part of the total capital budget.
3. State, and briefly describe, the so-called "problem areas" of capital planning.
4. "I decided to buy a new car because the ash trays were full." While one would not expect to take such a statement too seriously, one does often

hear it said that the car was turned in for a new model because it needed new tires. What replacement economics are involved in such a decision? If, strictly speaking, the only thing needed were a set of new tires, what decision would be called for if the urgency criterion were employed? Can you see any danger of "double counting" in this case?

5. State in your own words the meaning of "payout." Now translate your statement into a simple equation, defining each of the symbols used in that equation.
6. What is the relationship between payout, profitability, and project life?
7. Which is greater: after-tax payout or pre-tax payout? Can you state one in terms of the other, in an algebraic expression, assuming a 50 per cent tax rate?
8. State clearly the difference between payout and capital recovery period. Under what condition would they be equal?
9. Define rate of return on investment.
10. What is "annual cost"? What is its relation to capital recovery?
11. When a share of stock is purchased at $20, and carries a dividend of $1.00 per year, what are all the assumptions implicit in the statement that "the yield or return on the investment is 5 per cent?"
12. Prove that for very-long-lived projects (assuming a uniform annual cash flow and a single initial investment outlay) the return approximately equals the payout reciprocal.
13. Discuss the use of the payout reciprocal as an estimating tool in rate-of-return analysis. Under what conditions does it prove to be an excellent short cut? What circumstances would cause it to go far from the mark?
14. What are outlay-revenue ratios? In what method of analysis are they used? Why do they provide a suitable answer to the problems in which they are employed?

Chapter 11

CAPITAL MANAGEMENT (Continued)

The material in this chapter relates, in various ways, to the very important "cost of capital" concept. It is a concept which, either explicitly or implicitly, is woven throughout the subject of capital planning; and because of its great significance, we will examine it in most, if not all, of its ramifications. However, we shall do this in the natural course of discussing the remaining problem areas of capital planning as they were indicated in the preceding chapter. After that, we shall turn our attention to certain other aspects of the "cost of capital."

ESTABLISHING THE ACCEPTANCE CRITERION

The measurement of a project's rate of return,[1] or establishing its superiority or inferiority relative to other projects by whatever measure one might choose, is only one important step toward the construction of the final capital budget. Thus, having determined that the rate of return on a project is, say, 15 per cent, do we accept it or not? It would be manifestly imprudent to rely on some intuitive figure which "sounds" good or "seems" attractive, and so it becomes necessary to consider the establishment of some standard that will divide projects into two broad groups: those that are acceptable and those that are not.

Equilibrium of Supply and Demand

It was the great Alfred Marshall who depicted the forces of supply and demand as the two blades of the scissors, both of which were necessary for performing the function of determining equilibrium price. The analogy might be adapted to the problem under discussion. Thus, if the various proposed projects were arrayed in descending order according to their estimated rates of return, together with the dollar amounts of capital required by the respective projects, we would then have constructed what constitutes the firm's demand (schedule) for capital. The problem would then be to determine a capital supply schedule and, theoretically, the intersection point would indicate the desired volume of investment to be undertaken. All projects promising a return in excess of this intersection rate would be accepted; those with estimated rates less than the critical rate would be rejected.

---

[1] Because of its universal applicability to all types of projects we shall, for discussion purposes, use the *rate of return* as the appropriate measuring stick whenever referring to the evaluation of alternative investment proposals.

As nice and as neat as this approach seems to be, it is not readily useful for the problem at hand. The reason is that it is extremely difficult to establish a uniquely determinable capital supply schedule that will intersect the demand schedule at always the same point, which could then be accepted as the "cutoff rate." In other words, the supply of funds available (currently and potentially) to the firm is conditioned by a vast complex of factors: dividend policy, the firm's asset and liability structure, capital projects instituted in the past, current profitability, and many other factors which, in greater or lesser degree, are subsumed under the foregoing list. It is, therefore, an arbitrary oversimplification of the case to assume a given availability of capital (internal or otherwise) and to apply this (usually as a vertical supply curve) to the demand curve, with a resulting "equilibrium" cutoff rate. Since the firm's supply of liquid capital can be altered at will by changes in plans respecting debt retirement, dividend policy, working capital position, sales expansion, asset conversions, and so on, it is clearly necessary to exercise a great deal of precaution in designating the firm's capital availability.

The Potential Internal Supply of Capital. To any firm, the full potential internal supply of liquid capital consists of cash plus the funds that can be acquired by the sale of all other assets, less those conversions which result in the creation of accounts receivable. From this total must be deducted currently maturing cash obligations as representing a preemptive claim on the cash fund. As a practical matter, the internal cash supply to be made available for the acquisition of new assets is substantially less than that defined above, for it is necessary to reduce this sum by that portion which will be "reinvested" in existing assets—i.e., that sum represented by the existing assets which will not be sold.

The foregoing comments might well seem to the pragmatic decision maker to take on the appearance of a flight of fancy, for one is naturally inclined toward the attitude that a firm's operating assets are hardly ever seriously considered as a source of cash (except through their gradual conversion in the normal business process).[2] While this is generally true, it is so for reasons which do not necessarily negate the above considera-

[2] The economics of equipment replacement gives explicit recognition to salvage values as a part of replacement funds. Nevertheless, equipment replacement is a normal part of business operations, and what we have in mind above is a large-scale conversion of assets which have neither become worn out nor obsolete. And this would certainly be unusual.

The factoring of accounts receivable, particularly characteristic of the textile industry, is another example of operating assets which have come to be used as an immediate source of cash. But this, too, for that industry, is "normal."

tions. Theoretically, the decision to "reinvest" in any of the firm's existing assets should be taken only if the present value of their anticipated revenues exceeds the cash value of their current sale. Otherwise it would pay to exchange existing assets into others which, for the same investment, will produce a higher capitalized revenue stream. The fact that firms ordinarily do "reinvest" in existing assets might be interpreted as implicit recognition of this fact, though it is unlikely that specific thought along the lines suggested here are ever pursued, except in unusual circumstances.[3] In practice, however, an explicit calculation is usually not necessary because the assets in question are frequently too specialized (or involve dismantling costs or other substantial cash leakages) to provide, when sold, sufficient cash which, through investment in other assets, will enhance the firm's present value. That such considerations are nevertheless quite realistic, and may play a vital role in decision making, was explained in the previous chapter: the airline companies have for years relied on asset conversions as a major source of investible funds, and the capital gains from such sales have, in many instances, contributed the lion's share of net income. (This is, however, an extreme example of equipment replacement economy).

Before leaving this particular topic, but without becoming any more esoteric than has already been managed, let us be sure we have conveyed the opinion that, while it is probably a theoretically sound approach, the supply-and-demand technique is not a good workable method for solving the problem of determining an acceptance criterion. The method, it seems, is so capricious for this purpose that no two independent investigators are likely to arrive at the same solution for any given situation. In fact, it appears likely that no method can be expected to give precise and predictable results, but at least one of the techniques discussed below offers a promise of greater uniformity and accuracy.

Allotted Funds

Many firms meet the problem under discussion by allocating, either arbitrarily or by some simple "formula," a specified sum to be expended for capital purposes during a given fiscal period. Hence, the "allotted funds" method may embrace as wide a variety of techniques as there are firms that employ it.

To the extent that the allocations are arbitrarily determined, it is clear that a correct decision will be made only by coincidence, for no arbitrary allotment can possibly result in the establishment of an optimum cutoff point. And where allocations have been tied to some "formula," here too the optimum cutoff can be attained only by coincidence.

---

[3] In late 1957, and in 1958, the Penn Texas Corp. engaged in a liquidation of several divisions of the company in order to meet debts incurred in its unsuccessful attempt in 1957 to wrest control of Fairbanks Morse & Co. This is clearly a case of "unusual circumstances."

The major weakness of this approach is that the amount of funds allocated to the capital budget is predetermined—or tied to some consideration which has no direct connection with the contributions that the proposed projects are expected to make to the enterprise. Thus, many firms adopt some arbitrary dividend payout ratio (percentage of earnings paid out in dividends) and tie their capital expenditure programs strictly to the earnings that are retained, plus depreciation flows. It is typical to "earmark" depreciation funds for reinvestment, whatever the allocation method employed; and because of the inflation of the postwar years, most firms have reconciled themselves to requiring something in addition to depreciation funds to replace worn-out or obsolete facilities.

Debt Aversion. Many firms place an upper limit on capital expenditures in terms of most or even all of the cash earnings, the restriction being, in effect, that the firm avoid recourse to outside financing, especially with respect to the acquisition of debt funds. This aversion to debt has been true of many companies and is, at least to some extent, an outgrowth of the separation of ownership and control—a phenomenon discussed at length in Chapter 4. That such aversion does widely exist has been reported with remarkable consistency in many investigations of the subject, references to which are given in the bibliographic note at the end of this chapter.

The practical reasons for debt aversion seem to be based largely on the view that the greater profits engendered by trading on equity during favorable periods do not properly compensate management's risk exposure during times of economic reversal. In the latter case, management faces the possibility of stockholder revolt, or raids by outsiders—in either event, a threat to well-paying jobs. Since the executive's salary represents to him, in the typical case, a much more important source of income than his stock ownership—(many executives, in fact, own little or none of their company's stock, though the widespread use of stock options as a form of managerial compensation has done much to alter this situation) —it is quite understandable that "conservative" financial policies are so widely adopted.[4] Nevertheless, many industrial companies have engaged in debt financing to a degree which only a few years ago they would have considered to be excessive. This development is due to a combination of two things: a high level of prosperity that instilled an almost boundless optimism in many executives; and a high level of taxes that has placed a premium on debt financing.

Inverted Reasoning of Allotted Funds Method. In concluding this subsection on the allotted funds method of capital budgeting, it seems clear enough that what is wrong with any of its variations is that they all smack of "putting the cart before the horse." A rational approach requires that the volume of capital expenditures be determined by profit-making considerations, and not by any arbitrary methods which are at best only remotely related to considerations of profit. However, it must not be inferred that matters such as dividend policy, liquidity, capital structure, and the like should be ignored in executing the capital budget. It will be seen later in the chapter that such considerations have an appropriate place in the total picture.

---

[4] The recent challenge for the control of Montgomery Ward & Co. is an interesting example of exactly the opposite case. Here it was the management's ultraconservative policies with respect not only to finances but virtually to all the aspects of the company's operations that resulted in the well-publicized fight.

Cost of Capital

It would seem a perfectly logical principle, from the concepts of marginal cost and marginal revenue, that a firm would improve its earning power if the expected rate of return on a given project proposal exceeded the cost of acquiring the funds necessary for bringing the project into being. However, where there is widespread agreement about the theoretical meaning of the rate of return on a project (as defined earlier), the same cannot be said of the concept of cost as it relates to this particular problem. An immediate realization of the complexities involved is induced by questions such as the following: What is the cost, if anything, of retained earnings? How is the cost of new equity funds measured? How is the cost of funds determined when both debt and equity sources are tapped in financing capital expansion? In what way does dividend policy enter into the picture in its effect on capital costs, and if a firm pays no dividends does this mean that its equity funds are "free"? These and many other difficult questions must be settled in arriving at a satisfactory solution to the problem at hand.

Confusion Regarding the "Cost of Capital." The questions raised above relate to a concept that is discussed in the literature as "cost of capital," and it is fairly well accepted that the cost of capital to a firm provides the optimum cutoff rate or acceptance criterion for project proposals. The obstacle still to be surmounted, and a major one it is, is how to measure a firm's capital cost. The nature of the difficulty is indicated by the above questions: to take into account the different kinds of equity and debt capital the firm might rely upon, including such diverse sources as retained earnings, common equity, preferred equity, and all kinds of debt; and to "unify" or "subsume" them into some sort of over-all measurable concept which can be used as the firm's cost of capital or cutoff rate.

A significant part of the problem is that no general agreement about the precise meaning of the concept exists; yet, it is employed rather freely in the literature of capital theory. And while several economists have recently stepped out in the right direction,[5] some more notably than others, a workable solution is still wanting.

---

[5] See the bibliographic note at the end of the chapter.

Cost of Capital as an Opportunity Cost. It is generally accepted, and this is explicit in the leading books on the subject (i.e., Dean's and the Lutzes'), that a firm's cost of capital is an opportunity cost concept. The Lutzes go still further by making a distinction between what they call a "borrowing rate" and a "lending rate." The borrowing rate is conceived to be a rate at which the firm or entrepreneur can borrow, and, for any given firm (credit rating) at any given time the rate will vary with (increase with) the credit risk to which the lender is subjected. On the other hand, the entrepreneur has available to him the opportunity of investing any free funds he might have either in his own firm, or in some alternative outlet. The outside rate of interest available to him is the "lending rate" which, to him, is constant because of the substantially perfect competition that exists in the market for funds, and because his activity in the outside market is likely to have no more than an imperceptible effect.

This distinction between a borrowing rate and a lending rate points up one of the major sources of the confusion that has developed about the cost of capital concept. For purposes of fund raising, by whatever means, the borrowing rate is the appropriate measure of what the firm may be expected to pay for new capital; for purposes of discounting future cash flows to the present, the lending rate is the appropriate one because it reflects investment alternatives available to, and outside of, the firm. But the literature of capital theory and capital budgeting has used these concepts interchangeably as the firm's cost of capital, so that it is small wonder indeed that this area has remained so clouded and confused.

Structure of Rates. The confusion is further compounded by the fact that the borrowing and lending rates are themselves complexes, just as is "the interest rate"—a phrase so commonly employed in the economic literature. Actually, there exists not a single interest rate but a pattern or structure of interest rates. In the capital markets, a separate interest rate exists for every class of claims to be found there, varying as to maturity, issuer, protective convenants, nature and priority of rights, etc. These, together, comprise the complex structure of interest rates, and it is this complexity which is too often glibly glossed over in the frequent references to "the interest rate." Further, there also exists a huge and diverse volume of equities, on which no interest is paid, but which, by virtue of the dividends currently being paid on them or of the expectational receipts of dividends in the future, are tied to, or find their place in, the complex interest rate structure.

Much the same may be said of the "borrowing rate" and "the lending rate," each of which have been suggested as the measure of the firm's cost of capital. As a "borrowing" agent, or fund raiser, the firm has open to it a wide range of possibilities: various types of debt differing as to security, maturity, priority, etc., and equity, or common stock and various types of preferreds. (For the small firm, of course, it must be realized that the number of possible sources for new capital are greatly limited.) But what, then, is "the borrowing rate?" As a "lending" agent, or supplier of funds to others, an even wider range of possibilities exists, from riskless investments such as government bonds (of varying maturities and yields) to highly speculative equities. What, in this case, is "the lending rate?" In arriving at the answers to these two questions we will have achieved a solution to one of the most vexing problems in the theory and budgeting of capital expenditures.

The Earnings Yield

Internal and External Opportunities. The two questions raised above are actually tied up rather closely with one another, but we will approach the "lending" aspect first. In doing so it seems that it would be more meaningful to think of the function as that of investing rather than lending, since the investing function more clearly includes the acquisition of equities and titles, as well as claims. The rates of return on the investment opportunities confronting the entrepreneur outside the firm constitute, together with the returns on the investment opportunities within the firm, a vast array of alternatives available for selection. Because the many alternatives involve varying degrees of risk and nonmonetary advantages and disadvantages as well, it is not simply a matter of selecting those opportunities offering the greatest rate of return. It is, rather, necessary to recognize that the entire structure of rates represents a range of opportunities within which the firm itself may be fitted, and that the returns available on opportunities existing within the firm are strictly comparable only to those opportunities outside the firm which involve equivalent degrees of risk. Given sufficient freedom of flow of capital from industry to industry, opportunities of equivalent risk can be expected to provide comparable rates of return. Thus, the opportunity cost rate (outside investing rate) to be used is that which is available on investments equivalent in risk to that within the firm itself.

Expected Earnings Yield. In casting about for the correct measure of this opportunity cost rate, the most logical route leads to what might be called the "expected earnings yield" on the common stock. The reasoning goes as follows:

1. If the rate of return on a project exceeds the "cost" of the new funds needed to finance it, the future earnings available to the stockholders will be increased by accepting the project. And in terms of the profit-making goals of business firms, improvement in earnings available to the stockholders is a necessary condition for investment action.

2. The "cost" of the new funds is properly measured as a ratio of anticipated future earnings per share without the project to current net price per share. The latter is a measure of the dollars the firm can receive from the sale of stock (net of underwriting and flotation expenses); the former represents the expected productivity of existing capital in the firm. Thus, unless newly employed funds can produce incremental earn-

ings at the same rate as or better than the existing funds, the earnings available per share will decline.

The above "syllogism" is based on the assumption that the new funds will be acquired by sale of common stock, so that earnings dilution will result if the new funds are employed to finance projects which promise a rate of return less than that represented by the expected earnings yield on the existing stock. The argument as it stands, however, is incomplete for it must be made to encompass funds derived from other sources: retained earnings, sale of debt securities, and sale of preferred stock. And the solution will be meaningful only if we are able somehow to relate these diverse sources to one another. The basic element to which all these diverse sources and the costs of funds derived from them must be related is the opportunity cost rate defined earlier and described above as the "expected earnings yield." In expounding upon the relationships existing among the costs of the diverse capital sources that a firm might tap, we hope to make clear why we were justified in saying earlier that the "borrowing rate" and "lending rate" concepts of the Lutzes are inextricably intertwined.

Retained Earnings

If it were not for various leakages (the most important of which is the income tax) and other overriding considerations, all earnings could be distributed as dividends and, to the extent that expansion were deemed desirable by the directors, could be recouped by selling new stock in sufficient quantity to bring in the desired equity funds. This idealistic proposal has been made from time to time by various economists as a desirable means of turning over to the stockholders complete control of the corporate capital.

For various practical reasons, however, this would not be a very satisfactory procedure. Nevertheless, for the purposes of treating the problem under discussion, a 100 per cent distribution policy associated with the sale of whatever amount of stock is necessary for acquiring an adequate volume of funds to finance internal expansion does serve as a useful point of departure. Thus, considering only the leakages resulting from the income tax, and ignoring the other smaller leakages due to brokerage fees and underwriting expenses, a dollar of retained earnings is equivalent to two dollars of dividends paid to a stockholder in the 50 per cent tax bracket, since such a stockholder would be able to put back into the corporation (or into any other investment) only one dollar after paying his taxes.[6] What this means, then, is that if the opportunity-cost rate as measured by the expected earnings yield is, say, 12 per cent (so that projects promising less than this should be rejected if they are to be financed by sale of new stock), it would be to the economic advantage of the stockholders to accept projects promising substantially less than 12 per cent if financing is to be done from retained earnings. How much less than 12 per cent, i.e., how much below the earnings yield, it would be desirable to go, would depend on the tax brackets of the corporation's stockholders.

---

[6] We are ignoring the fact that the tax laws exclude a small portion of dividends from taxes and provide a tax credit for the balance. To the small stockholder this effectively eliminates or diminishes the tax he must pay on his dividend income. However, the existence of these tax laws only requires that we adjust our conclusions accordingly; the laws do not invalidate the general argument itself.

In the close-held corporation, where the insiders own as well as control the enterprise, the tax situation of the managerial team can be readily enough applied as a guide to policy. But in the publicly held corporation where the tax brackets are likely to range from zero for the tax-exempt institutional stockholder to extremely high percentages for the very wealthy individual, the problem is not so easily solved. Perhaps a workable solution would be to assume a median tax bracket, or something above that based on the logic that the lower-income groups in the economy do not own stock (at least not to a degree sufficient to warrant undue concern on the part of the management). If the 30 per cent tax bracket were assumed to be the appropriate guide to employ, it would mean that a project need promise only 70 per cent as much a return when financed by retained earnings. Hence, a cutoff rate of 12 per cent applied to projects to be financed by sale of new stock would be equivalent to a rate of 8.4 per cent where retained earnings are to be used.

Sale of Senior Securities

The use of bonds and preferred stock as media for acquiring long-term capital is referred to as *trading on equity*—a concept which is generally well enough explained in elementary textbooks in finance. The general principle involved is that the profits available to the stockholders will be increased if the rate paid on the senior capital proves to be less than the rate earned on that capital when employed in the firm. This advantage is, unhappily, offset by the risk exposure resulting from trading on equity, for the profits available to the owners will be less, and the losses greater, if the rate earned on the senior capital should prove to be less than its cost.

How can we reconcile this with the opportunity cost rate defined earlier in terms of earnings yield? From what we know about the costs of debt funds, it would appear that the cutoff rate expressed in terms of the earnings yield is much higher than debt capital costs, so that the principle of applying such a rate for cutoff purposes runs quite contrary to the suggestions flowing from the trading-on-equity principle. Thus, a firm whose opportunity cost rate is, say, 15 per cent would apparently be missing many profitable opportunities by rejecting projects promising less than that if it can finance them by borrowing at, say, 5 per cent. This simple logic seems reasonable enough, and is the sort usually employed in the typical textbook treatment of trading on equity.

Equally simple and reasonable logic, however, indicates that the above conclusion is erroneous. Can a firm that has shown vigorous growth —such as Minnesota Mining & Manufacturing, or International Business Machines—continue its exceptional rate of growth by accepting projects with low earning power simply because the funds for financing such projects are available to it at or slightly above the prime rate? Obviously not! And if it did follow such a course, the very high price-earnings ratio (low earnings yield) would fall (rise) precipitously, reflecting a sharp increase in the cost of any new equity funds it might later seek. In a less spectacular fashion the same would happen to any stock, though the reasons may not be quite so apparent. There are actually many subtleties involved which are typically ignored (often unconsciously) in the elementary textbook treatment of trading on equity, for the ramifications that result from combined debt and equity financing are very complex.

It is generally quite well understood that when a firm engages in debt financing it exposes itself to risks which, once the debt begins to approach a rather sizeable amount relative to the total capital structure, increase in rapid geometric fashion compared to the increase in the debt itself. This reflects itself, of course, in the leveraged effect on corporate earnings; but it reflects itself also in the earnings yield (price-earnings ratio) of the stock in adverse fashion. The increase in the cost of new equity funds to the firm is directly traceable to the debt financing and therefore must be properly considered as part of the real cost of borrowing. It has been suggested, in fact, because of this interplay of forces between debt and equity financing, that the true measure of all financing costs be taken to be the cost of equity funds. The reasoning involved is that managements recognize the hidden costs of borrowing—in the form of the risks of default mentioned above, as well as the loss of managerial flexibility and freedom of action in the form of dividend restrictions, working capital requirements, and other constraints imposed by the bond indenture. Thus, given the firm's capital structure, management will undertake new financing in that medium (debt or equity) which is least costly, so that there will exist, or tend to exist, an equality between the (real) cost of borrowing and the cost of equity. Because of this equality, the best way to measure cost is to look at the earnings yield on the stock, for this is objectively measurable, whereas most of the "real" borrowing costs are subjective and not readily measurable (default risks and constraints).

The Optimum Capital Structure Approach

The above argument is actually quite ingenious, and has a great deal to commend it. Most important is that it recognizes the significance of the interaction of debt and equity. But because much of the real costs of borrowing are subjective, it means that what might be a moderate cost to one management is excessively burdensome to another. The result is that it becomes impossible, in the first place, to state objectively the combination of debt and equity that must be sought in striving for the point at which the marginal real cost of debt financing equals the marginal cost of equity financing; and, in the second place, it would seem to follow from the argument presented above that management always strives to achieve this balance anyway, and in its own way, so that whatever the capital structure might be it is always at, or tending toward, an optimum combination.

The Significance of the Firm's Capital Structure. The last conclusion has an almost teleologic quality and, for that reason at least, is suspect. It is, of course, a principle or precept that could well become a part of the normative structure of neoclassical economics; but, as it stands, the equation of marginal borrowing and equity costs in the individual firm is presumed *somehow* to occur, and this is neither probable nor useful as a guide to the management that would like to have some concrete suggestions on how to achieve this goal. A solution does exist, however, and that is to accept again the impersonal dictates of the market.

As indicated earlier, the market reflects the cost of borrowing in the valuation which it places on the common stock of the company. But borrowing need not necessarily involve positive real costs over and above the nominal out-of-pocket interest charges. If the borrowed funds are expected to produce earnings at a rate that is at least equal to the earnings yield, and if the total debt is a very small portion of the capital structure, the market might welcome the decision to borrow by reducing the rate at which the firm's earnings are capitalized. Thus the real cost of borrowing might actually be less than the interest charge. This implies that there is, in the opinion of the market, an optimum capital structure which reflects the firm's desire to take advantage of the benefits provided by leveraging, while at the same time keeping the risk of default under control.

Variability of Optimum Structure. The optimum ratio of debt to equity in the capital structure will vary considerably from one industry to another and, to a significant extent, among companies within a given industry. This variability enormously complicates the problem of establishing criteria to serve as guides in constructing optimum financial structures. As a result it is necessary to settle for general principles rather than rules of precision, but even general principles will help greatly to fill the void that currently exists.

We might well approach the problem by asking what it is that the earnings yield is supposed to reflect. As a relationship between average future annual earnings and recent average price, the earnings yield actually reflects a host of factors, some of them imponderable. However, it may be reasonably assumed that in the long run all significant factors, whether tangible or not, will reflect themselves in the record of earnings and dividends. Hence, it may be argued that it is unnecessary to attempt to evaluate such difficult elements as managerial ability, personnel relations, competitive position, new product development, operating efficiency, and so on. These are all important considerations and they help to explain why the earnings pattern is what it is. But they are important also for the effect they are expected to have on future earnings and dividends. Hence, these basic series provide us with the materials we need.

The next step is to fashion these materials into the forms in which they are used by investors. Since it is the market's attitudes and psychology that determine the cost to the firm seeking new capital, it is upon these attitudes that a meaningful and measurable cost of capital concept must be built.

Defining Anew the Cost of Capital. From what has been said above, our earlier definition of the cost of capital simply in terms of earnings yield is not suitable. Rather, a more fruitful approach would be to define the "cost of capital" as the cost of equity funds (still measured by earnings yield) *when the firm has what the market considers to be a well-balanced capital structure.* For with such a structure it may be assumed that the marginal real cost of borrowing equals the marginal cost of equity funds. If a firm is considered to be excessively leveraged, the marginal real cost of borrowing (because of the risks involved) will exceed the marginal cost of equity financing; where a firm is too conservatively capitalized, additional borrowing should involve lower costs than would be required on equity funds. It may further be pointed out that each new financing, whether by debt or by equity means, can be expected to affect the marginal cost of financing by the alternative method. For example, a firm that is already believed to be too highly leveraged would cause the marginal cost of equity funds to increase if it were to engage in additional borrowing; on the other hand, if the same firm were instead to employ equity financing, this would reduce the marginal real cost of borrowing.

Some Unsolved Problems

The problems that remain to be solved, with respect to the actual determination of a firm's cost of capital, taking into consideration the impact of the firm's capital structure on its capital costs, are: (1) the establishment of standards of optimum capital structures; and (2) a means of relating the variations of those structures which depart from the optimum to the cost of that particular firm's capital.

If it were necessary, for practical applications, to set down a precise system of standards and procedures, we would have before us an impossible task. As it is, the problem is a difficult one, but it can be coped with. Our concern is only that we arrive at a useful and usable measure of the cost of capital, and this requires, from what we have said thus far, that we recognize, and try to adapt to, the tastes and attitudes of the market. While a precise measure of these tastes and attitudes would be desirable, it is by no means necessary for our adopting a meaningful course of action.

Investment Ratings. Actually, standards exist which have been employed for many years by the investment community and which can be adapted to the problem being presently considered. We refer to those standards that have been employed to distinguish among different qualities of stocks and bonds. It is the practice in security analysis to rate both stocks and bonds for their investment quality. These ratings are particularly important with respect to bonds, for the investment status given to debt securities by the leading rating services (Moody's, Standard and Poor's, and Fitch) will determine whether or not they will find their way into the portfolios of institutional investors. This means, as well, that the yield basis on which the bonds will be taken up will be accordingly affected—in fact, the ratings on these securities will even affect the terms under which subsequent sallies into the capital markets could be made. This applies to a firm's equity financing too, though perhaps in somewhat lesser degree.

These investment ratings thus have a clear connection with a firm's cost of capital, in that they both reflect and influence the firm's credit standing in the capital markets. However, the security ratings are based on statistics and qualities which have, in many cases, no direct connection with the composition of the firm's capital structure. For the purpose of coping with this particular problem, it is necessary to narrow down to those factors which relate to what the market would consider to be an optimum capital structure, and the way in which these considerations affect the security ratings and the earnings yield on the common stock. In doing so it is necessary to classify companies into various groupings. Thus, an electric power company would obviously have a different optimum structure than a steel warehousing firm; the former would be expected to be much more highly leveraged. In fact, an electric power company that had only 30 per cent of its capital structure represented by debt and preferred stock would be exceedingly conservative, and would be considered so to the detriment of the common stockholders. Such a company would obviously have been financing much of its growth via sale of common stock, the result of which would have meant repeated dilution of common stock earnings. This would further mean that the company's growth trend was being unnecessarily dampened and it would result in the capitalization of the company's earnings, by the market, at a higher rate than necessary, i.e., a higher cost of capital for the firm. We have here not only an example of how the composition of a firm's capital can affect its earnings record and, through it, its cost of capital, but an illustration as well of what was meant earlier when it was stated that all factors—tangible or intangible, quantitative or qualitative—will in time reflect themselves in the earnings and dividend records of the company.

The typical classification for security investment purposes is the quadripartite grouping of public utility, railroad, financial, and industrial. This classification is unsatisfactory for our purposes, however, because of the heterogeneity of the individual groups—especially the industrial category. For example, the dairy, food chain, and tobacco companies have better records of earning and dividend stability than most railroads and even some public utilities. Thus, the quadripartite arrangement serves only as a useful point of departure: it is a recognition in the first place that the market seeks to establish capitalization rate differentials in terms of stability of earnings and dividends, and that this is a reflection, to a large extent, of the nature of the business or industry in which the firm competes. This is actually another way in which the uncertainty factor enters into the picture.

STRICTLY BUSINESS by McFeatters
ANNUAL CONVENTION
Dale McFeatters
Reg. U.S. Pat Office Publishers Syndicate

"J. B.'s a man of good stock—mostly Standard Oil and General Motors!"

The Omnipresence of Uncertainty

One might well argue that uncertainty is the only factor that explains the discrepancies in the earnings yields existing in the market at any given time. It is very common, however, for investment analysts to discuss growth as a factor distinct from investment risk, and to explain extremely low earnings yields (as measured in terms of current price and latest twelve months' earnings) in terms of the growth potential anticipated by the market. Thus, International Business Machines common stock sold, during the 1956–57 bull market, at a current earnings yield of between 2 and 2.5 per cent at a time when intermediate-term government bonds were yielding close to 4 per cent (the dividend yield on IBM was under 1 per cent). It would obviously be false to conclude that IBM was considered to be "safer" than government bonds. The answer lies rather in the fact that the investor in either case is actually buying a stream of future income, and that, for the government bonds, the future stream was to flow at a constant rate, while it was expected that the stream produced by the IBM investment would flow at a rapidly expanding rate. For this reason, the earnings yield as a measure of cost of capital must be defined, and has so been defined here, in terms of future rather than current annual earnings.

While the growth factor seems to explain certain discrepancies in current earnings yields, uncertainty here too plays an important part. For there is uncertainty concerning: (1) whether a given investment will produce a *growing* stream of earnings; (2) the *rate* of growth that will be achieved; and (3) the general level of earnings and dividend yields prevailing in the market in the future, which will, of course, affect the liquidating value of the investment at that time. If the market is performing its valuation function properly, all of these factors are reflected in the current earnings yields of common stocks.

The Need for Continued Research

The foregoing considerations point out the route along which a fruitful investigative research of the problem might proceed, but, at this point, we are not as yet able to make too definitive a statement. There is reason to believe, for example, that an electric power company can safely carry a capital structure represented by less than 30 per cent in common stock, and that 50 per cent would probably be too conservative.[7] On the other hand, a cyclical business would possibly demand, in the interest of safety, little or no debt and, perhaps, up to a maximum of 25 per cent of preferred stock if leveraging is to be employed at all.

Exactly what the optimum capital structure should be for any given type of business, i.e., what the capital market seems to view as an optimum structure, remains to be determined, and would make a very interesting and useful study. It would provide the basis for making fairly accurate estimates of a company's cost of capital. While rather crude approximations are currently possible for any given industry—investment attitudes are always making themselves felt—such estimates would be greatly improved by a careful research survey. And once it is possible to accept confidently a given capital structure as the optimum for a certain type of company, that company can: (1) gradually adjust its structure toward that optimum, thereby improving (reducing) its cost of capital; and (2) estimate fairly accurately what its cost of capital is and use it as its cutoff rate for its current capital planning. Until this "scientific" approach becomes possible, business managements will simply have to content themselves with less reliable estimates.

---

[7] See Table 11–4, p. 422, and the discussion of the data contained therein.

CAPITAL COST PATTERNS

The foregoing treatment of the cost of capital is important in providing the basis for managerial decision making in the area of capital planning. Also of great importance, to both the academician and the businessman, are the trends and patterns revealed in the available capital data. This section is devoted to a presentation and discussion of some of these data.

Explanation of Table 11–1

In accordance with our definition of cost of equity capital, Table 11–1 presents data going back as far as 1920. Equity capital costs are shown as they have been reflected by the Dow-Jones Industrial Average. These costs are computed as earnings yield figures (determined as a ratio of earnings to the mean annual price), and appear in the table as the post-tax cost of equity capital, because the earnings figures shown are net of taxes.[8] To convert these to a pre-tax basis, it is necessary to apply the effective tax rate (including normal and surtax). These rates are shown in a separate column in Table 11–1.

The pre-tax cost of equity capital can thus be compared to the cost of debt capital, the latter being represented by the yield on Standard & Poor's Composite Index of high grade corporate bonds. No adjustment for taxes is necessary in these figures since interest is charged before income taxes are imposed. The difference between the pre-tax cost of equity capital and the (pre-tax) cost of debt capital is then shown in the last column as the "differential cost." There is no significance to the choice of this term: it is simply meant to represent the difference between the costs of equity and debt capital.

To one accustomed to thinking of capital costs in terms of interest charges and bond yields, or even in terms of dividend yields, the data

---

[8] While we have not taken the trouble to do so, a much more meaningful approach to ascertaining equity capital costs, historically, would be to develop a so-called "normal" earnings-yield ratio, either as a moving average around a long-term trend, or as an average relationship between earnings and prices as determined from "average" or "normal" years. Merely using an annual mean price is not entirely satisfactory, although for our purposes this simplified approach is good enough.

TABLE 11-1

COSTS OF CAPITAL AS REFLECTED IN THE DOW-JONES INDUSTRIAL AVERAGE AND STANDARD & POOR'S COMPOSITE INDEX OF HIGH GRADE CORPORATE BONDS

YEAR	DOW-JONES INDUSTRIALS		CORPORATE TAX RATE[2]	COST OF EQUITY		COST OF DEBT	DIFFERENTIAL COST[5]
	Mean Price[1]	Earnings		Post-tax[3]	Pre-tax[4]		
1956	491.70	$33.34	52%	6.78%	14.12%	3.40%	10.72%
1955	438.30	35.78	52	8.16	17.00	3.04	13.96
1954	342.13	28.40	52	8.30	17.29	2.87	14.42
1953	274.64	27.23	52	9.91	20.65	3.18	17.47
1952	274.17	24.76	52	9.03	18.81	2.95	15.86
1951	257.68	26.59	50.75	10.32	20.95	2.84	18.11
1950	216.14	30.70	44	14.20	25.36	2.59	22.77
1949	181.06	23.54	38	13.00	20.97	2.64	18.33
1948	179.27	23.07	38	12.87	20.76	2.80	17.96
1947	175.03	18.80	38	10.74	17.32	2.58	14.74
1946	187.81	13.63	38	7.26	11.71	2.51	9.20
1945	173.58	10.56	40	6.08	10.13	2.61	7.52
1944	143.37	10.07	40	7.02	11.70	2.78	8.92
1943	132.54	9.74	40	7.35	12.25	2.80	9.45
1942	106.31	9.22	40	8.67	14.45	2.85	11.60
1941	119.97	11.64	31	9.70	14.06	2.84	11.22
1940	132.32	10.92	24	8.25	10.86	2.92	7.94
1939	138.68	9.11	19	6.57	8.11	3.02	5.09
1938	128.68	6.01	19	4.67	5.77	3.20	2.57
1937	154.02	11.49	15	7.46	8.78	3.30	5.48
1936	164.00	10.07	15	6.14	7.22	3.34	3.88
1935	122.58	6.34	13.75	5.17	5.99	3.61	2.38
1934	98.12	3.91	13.75	3.98	4.61	4.14	0.47
1933	79.41	2.11	13.75	2.66	3.08	4.69	−1.61
1932	65.00	−0.51	13.75	—	—	5.28	—
1931	134.07	4.09	12	3.05	3.47	4.55	−1.08
1930	225.79	11.02	12	4.88	5.55	4.71	0.84
1929	289.93	19.94	11	6.88	7.73	4.86	2.87
1928	245.66	15.97	12	6.50	7.39	4.63	2.76
1927	177.56	8.72	13.5	4.91	5.68	4.65	1.03
1926	150.92	11.39	13.5	7.55	8.73	4.77	3.96
1925	137.19	13.89	13	10.12	11.63	4.93	6.70
1924	104.42	10.88	12.5	10.42	11.91	5.07	6.84
1923	95.57	8.25	12.5	8.63	9.86	5.22	4.64
1922	91.01	9.11	12.5	10.00	11.43	5.17	6.26
1921	72.70	2.10	10	2.89	3.21	6.03	−2.82
1920	88.31	9.12	10	10.33	11.48	6.18	5.30

[1] Annual mean price of high and low values.

[2] Combined normal and surtax rates.

[3] Earnings yield ratio of earnings to mean price.

[4] Adjusted earnings yield post-tax earnings yield divided by 100% minus corporate tax rate.

[5] Pre-tax earnings yield minus cost of debt.

Sources Barron's Publishing Co , Inc , *The Dow-Jones Averages. Commercial and Financial Chronical* (July 11, 1946), p. 1 ff. The Tax Foundation. *Facts and Figures on Government Finance.*

appearing in the pre-tax equity cost column undoubtedly seem shockingly high. Among the more interesting patterns discernible in the table is the sharply contrasting downtrend in the cost of debt capital from the period of the 'twenties to the decades of the 'forties and 'fifties, as compared with the uptrend in pre-tax equity capital costs during the same span of time.

The data of Table 11–1 are regrouped into five-year intervals and presented as averages in Table 11–2. The purpose for doing this is to reduce short-term fluctuations in the data, and to highlight, for those who are unable to discern it for themselves, the basic shifts that have taken place during the entire period covered.

Determinants of Cost Patterns

The discussion of the pattern of interest rates earlier in this chapter applies in full to the determinants of the pattern of capital costs. It is clear that the two concepts are closely related, if not more or less synonymous, when we recognize that bond yields such as those shown for "cost of debt" in Tables 11–1 and 11–2 are part of the structure of interest rates, so that the forces which determine bond yields (as costs of debt capital) also determine the place of these yields within the interest rate structure. However, capital costs are determined by a host of forces that exert their influence on the capital markets, and since these forces do not all work in the same direction, it is not always an easy matter to explain the net effects, let alone to predict the direction of new and longer-term changes.

The Federal Authorities. The single most important determinant of the interest rate structure (and therefore of the cost of debt capital) is the federal government, making itself felt through the operations of the Treasury Department and the policies of the Federal Reserve authorities. While matters pertaining to money and credit are officially the province of the "Fed," the Treasury's operations are much too important for its effects on the capital markets to be ignored. The Treasury enters the money market each Monday with an offer of 91-day bills which, during 1957, ranged in the neighborhood of about $1.75 billion. Generally, this offering is simply a refunding of a similar volume of maturing bills, but even when this is the case the Treasury must still set a rate that will: (1) be taken by the market without shaking it unduly; and (2) be in general line with the objectives of the Federal Reserve authorities. In addition to these weekly bill offerings, the Treasury enters the market less regularly with longer maturities for sale. These must not only be fitted into the existing structure and be reasonably consistent with monetary objectives, but must also suit the Treasury's own purposes: (1) provide the desired volume of funds; and (2) lengthen or shorten the average debt as desired by the Treasury's policy-making officials.

The Fed, in whom is vested the responsibility for pursuing monetary and credit policies designed to promote economic growth and sta-

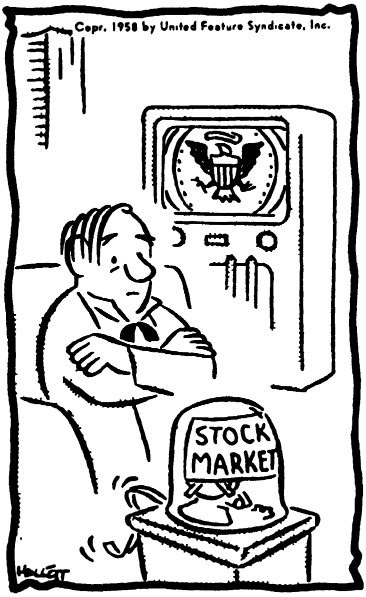
TABLE 11–2

FIVE-YEAR AVERAGE COSTS OF CAPITAL
(As derived from data of Table 11–1)

PERIOD	DOW-JONES INDUSTRIALS		CORPORATE TAX RATE	COST OF EQUITY		COST OF DEBT	DIFFERENTIAL COST
	Mean Price	Earnings		Post-tax	Pre-tax		
1952–56	364.19	$29.90	52.00%	8.21%	17.10%	3.09%	14.01%
1950–54	272.83	27.54	50.15	10.09	20.24	2.89	17.35
1945–49	179.35	17.92	38.40	9.99	16.22	2.63	13.59
1940–44	126.90	10.32	35.00	8.13	12.51	2.84	9.67
1935–39	141.59	8.60	16.35	6.07	7.26	3.29	3.97
1930–34	120.48	4.12	13.05	3.42	3.93	4.67	−0.74
1925–29	200.25	13.98	15.20	6.98	8.23	4.77	3.46
1920–24	90.40	7.89	11.50	8.73	9.86	5.53	4.33

bility, actually "sets" the pattern of rates which prevail in the market. Through the exercise of controls over the rediscount rate, member bank reserve requirements, and selective credit controls (such as margin requirements for stock purchases) the Fed produces a marked effect on the general level of the rate structure. On the other hand, its open-market operations, a tool that was originally supposed to enable the Reserve authorities to "make the rediscount rate effective," is capable of shaping the rate structure itself. This is possible by concentrating market operations on the shorter or longer end of the maturity scale, as the case might be.

Merry Washington


Copr. 1958 by United Feature Syndicate, Inc.
STOCK MARKET

"And now, the President of the United States . . ."

While other forces, discussed below, have important effects on the general structure of rates, the government bond market is almost entirely subject to the operations of what amounts to the monolithic force of the Federal authorities. One might argue that the manner and direction in which this force is applied is not entirely self-willed, but must suit the more basic dictates of economic conditions and objectives. Thus, Treasury operations depend on fiscal requirements and Congressional appropriations; the Fed's controls are relaxed or more stringently imposed depending on whether conditions call for credit relaxation or restriction. Be this as it may, the Federal force does not automatically arise out of market conditions themselves, but is imposed upon the market as a result of positively determined policy.

Public Psychology. For lack of space we will here interpret "public psychology" very broadly and treat it rather briefly. It includes such elements as "consumer optimism," "investor confidence," and "business outlook." These frequently point in the same direction at the same time, and their effects are most importantly felt on the yields which prevail on security issues other than Treasury debt obligations. Thus, while it is true that the Federal authorities have an enormous effect on all yields by setting the pattern through their government bond operations, the question of how, and to what degree, other yields will respond depends on "public psychology."

Except for the so-called "money-rate bonds" (issues which are of such high quality that the possibility of default is virtually nonexistent, so that their prices fluctuate strictly in line with government bonds), corporate bonds and common and preferred stocks will move in varying sympathy with the yields available on government bonds. Thus, whether the spread between bond and stock yields will be narrow or wide, or whether it will favor one type of investment as against another, depends more on the factors subsumed under "public psychology" than it does directly on the Fed's credit policies. For example, one would expect that dividend yields on good quality stocks would "normally" exceed the yields available on high-grade corporate bonds. This has happened more often than not (in varying degree), but it was not true from 1921 to 1929, and in 1933 and 1934. The reasons for such departures from "normal" are rarely the same, although certain interesting aspects are worth pointing out. Thus, one is likely to find bond yields to exceed or be only slightly less than stock yields during booming prosperity or deep depression. In the latter situation stock yields decline sharply as dividends are cut or omitted, while the fixed interest charges on high-grade bonds usually continue to be paid. During sanguine periods of prosperity, buoyant optimism leads investors to project unlimited economic growth for the privilege of which they are willing to buy equities at prices that provide absurdly low income and earnings yields. Table 11-3 presents a few selected years to show the sharp contrasts that are possible.

TABLE 11-3

COMMON STOCK YIELDS VERSUS BOND YIELDS AS REFLECTED BY THE DOW-JONES INDUSTRIAL AVERAGE AND STANDARD & POOR'S COMPOSITE INDEX OF HIGH-GRADE CORPORATE BONDS, SELECTED YEARS

Year	Stock Yields	Bond Yields	Differences
1956..........	4.68%	3.40%	1.28%
1950..........	7.46	2.59	4.87
1945..........	3.85	2.61	1.24
1934..........	3.73	4.14	−0.41
1927..........	3.40	4.65	−1.25

Source: Based on data in Table 11-1.

Individual Factors. Thus far we have considered the market forces which affect the general level and pattern of rates. Within this pattern must be fitted each individual security (investment), and the forces that have been discussed above apply also to the individual issue. But, in addition to those general forces, the individual investment is subject to specific forces. Thus, the decision to raise the tariff on lead imports, or to reduce further the allowable days of crude oil production in Texas, can be expected to affect specific industries or companies more than others, although it is conceivable that their effects might be traceable to many other segments of the economy. Similarly, if the outlook for automobile or housing demand is unpromising, the yields on the securities of the companies involved will reflect this fact. In short, many of the factors discussed above as part of "public psychology" can be found to apply, at times, with particularly telling force to specific companies rather than to industry in general. All of these factors, both those of general and specific import, can, in fact be classed under the heading of that all-embracing term "uncertainty." We are then able to say, simply, that yields on treasury securities are lowest because there is no question of default involved; that money-rate bonds offer somewhat higher yields; and so on down the scale to low-grade equities on which dividend and earnings yields are large.[9]

Corporate Income Taxes. The single most important determinant of the cost of equity capital has been the corporate income tax. To verify this fact one need only look back at Table 11–1 and note the almost perfect correlation between the corporate tax rate and the pre-tax cost of equity capital. This force might have been discussed above as one of the Federal powers, or even as one of the general factors that help mold the economic environment in which firms operate. But so important has this single factor become in the picture of equity financing that it well warrants special treatment.

It is clear from the above discussion that the forces which determine the *post-tax* cost of equity capital are many, and we can hardly claim to have delved deeply into the subject in these few pages. Yet, the effect of the 52 per cent corporate tax rate is, by itself, greater than all of the other factors combined in determining *pre-tax* equity costs. Compare, for example, the years 1920 and 1951. Post-tax equity costs were almost identical. But in 1920, when the corporate tax rate was only 10 per cent, the pre-tax cost of equity capital was barely larger than the post-tax cost; in 1951, with a corporate tax rate of just under 51 per cent, the pre-tax cost was more than twice the post-tax cost. The trend in corporate income taxes being what it has been, particularly since World War II, it is no wonder that reference is frequently made to the fact that our tax policies have placed debt financing in a favorable position.[10]

[9] Current dividends and earnings might be zero but the possibility exists, reasonable or not, that future earnings will permit payments of such order as to make the current price attractive, i.e., at least in line with prices of other securities.

[10] These advantages have been variously reflected in postwar corporate financing. For example, the subordinated debenture is an invention to take advantage of the

Of course, where post-tax costs are relatively low, as has been the case from time to time (see Table 11–1), even a 52 per cent corporate tax does not produce pre-tax costs of great enormity: doubling a small number results at most in something which is itself not much more than small or moderate. However, compared with the costs of debt capital, post-tax equity costs have usually not been particularly small, and a doubling of the more typical post-tax equity costs leads to pre-tax costs of rather shocking size. The implications for managerial decision making, discussed below, are of vital significance to all business firms.

CORPORATE CAPITAL STRUCTURES

The postwar boom was occasioned by the problem of turning out sufficient goods to satisfy the great pent-up demand of the consuming public. Thus, where the major problem had previously been how to move the goods that were being produced (selling problem), the postwar period was characterized by the problem of how to finance the capacity required to furnish the goods being demanded.

Postwar Corporate Conservatism

Plant expansion during the postwar period has been financed primarily from internal sources—retained earnings, depreciation, depletion, and amortization. However, while the capital markets have played a smaller role, relatively, during the recent postwar expansion as compared with the decade of the 'twenties, their importance in the capital formation that has been taking place can hardly be ignored. Thus, during the ten-year period from 1946 through 1955, the capital markets supplied approximately $88 billion to nonfinancial corporations (including the sale of stocks and bonds, term bank loans, and long-term borrowings from insurance companies and other suppliers of long-term credit). This compared with an internal cash flow (retained earnings, depreciation, etc.) of about $170 billion. Thus, one third of the permanent capital of such corporations came from external sources, as contrasted with two fifths during the ten-year period 1920–1929.

Actually, the conservatism of industrial corporations can be further emphasized by pointing out that the lion's share of the external capital financing in the postwar era was effected by the utility companies alone. These accounted for approximately half of the debt financing and about four fifths of the equity financing. Further, while corporate debt has risen consistently since 1939, stockholders' equity has risen much more rapidly, as has the ability to service the debt (corporate income relative to interest charges and sinking fund requirements). For example, where

---

fact that interest on debt is deducted before the income tax is determined. In the prewar days of lower taxes, these subordinated debentures would almost certainly have been preferred stock.

interest charges absorbed about one third of corporate gross income in 1929, they expropriated only one tenth of the gross in 1955.

Corporate Borrowing Policies

The great cost advantages of debt capital as against equity capital have already been indicated earlier, so that it is not surprising to note a slow awakening to this fact on the part of many corporate managements that heretofore would not have considered borrowing except for emergency purposes. In some cases, borrowing has been born out of the necessity of financing rapid expansion. Thus, companies that have been historically opposed to debt have grudgingly accepted the need for debt capital to the extent, at most, that cash flows from depreciation and amortization would be sufficient to meet interest charges and sinking fund requirements. As a group, the chemical companies provide an example of a fast-growing industry historically free of debt (more or less), which has had to accept it as, perhaps in some cases anyway, a necessary evil. Dow Chemical and Union Carbide have actually decided to secure whatever benefits debt financing has to offer by leveraging their structures substantially. (Dow, at its peak in 1952, had a structure comprising 50 per cent debt but has since reduced it to 24 per cent through conversions and build-up of equity; Union Carbide's debt is one third of the capital structure.) On the other hand, du Pont, the largest in the industry, has steadfastly stayed clear of debt, adhering to the dictates of its founder, 150 years ago, that this was the first principle of successful financial policy.

Even more startling contrasts are available in the nonferrous metals companies. The aluminum industry has undergone a phenomenal expansion in capacity during the postwar period and the companies involved have had to accept debt capital to bring their large and numerous projects to fruition. The result is that Aluminium, Ltd. shows a huge 49 per cent debt element and 8 per cent preferred in its structure, Aluminum Company of America has 38 per cent in debt and 6 per cent in preferred, Kaiser Aluminum & Chemical exhibits 55 per cent in debt and 15 per cent in preferred, and Reynolds Metals carries 60 per cent in debt and 6 per cent in preferred. Contrast this with the capital structures of such leading metals companies as Kennecott Copper, Phelps Dodge, New Jersey Zinc, and International Nickel with "clean" balance sheets—100 per cent common equity. Capital expenditures by these companies have been "more orderly"—tied to available funds. The difference, however, has not been so much a contrast in managerial philosophy as it is a difference in underlying growth in demand for the respective products and the need to put in place the capacities to meet these demands.

Endless examples of interindustry and intraindustry contrasts are possible. Table 11–4 presents the capital structures of a few selected industries, and it is quite clear, as one would expect, that the electric power companies stand out as strikingly different. One might say that during the last twenty-five years they have been the "pacesetters" or "innovators" in the field of finance. If this is so, it is because their regulated operation, coupled with the fast-growing demands on electric power, has made finance a more important field of activity to them than it has been for the typically successful industrial company which could so readily rely on retained earnings for financing its capital program. Thus, the common equity portion of the electric power structures accounts for 38 per cent of the total as compared with generally twice that amount for the other industries shown in Table 11-4. Still, within the industry, we find a divergence of managerial philosophy reflected in structures, with common equities of 54 per cent in Puget Sound Power & Light, 50 per cent in Boston Edison, and 49 per cent in Community Public Service, down to 32 per cent in Atlantic City Electric, 31 per cent in Gulf States Utilities, and 28 per cent in Pennsylvania Power & Light. It is further clear from the table that what is a very conservatively capitalized company (such as Puget Sound Power & Light) for the electric utility industry would be "highly leveraged" for even such a stable industry as foods.

TABLE 11-4

CAPITAL STRUCTURES, NET PROFITS, DIVIDENDS, AND PAYOUT RATIOS OF SELECTED INDUSTRIES IN 1955
(In millions of dollars)

	Petroleum (23 Corps.)	Chemical (32 Corps.)	Food (30 Corps.)	Nonferrous Metals (14 Corps.)	Electric Power (35 Corps.)
Long-term debt*	$ 2,420	$1,429	$ 596	$ 735	$7,960
Preferred stock	120	458	249	214	2,007
Common stock	5,959	1,213	840	1,059	4,157
Surplus	10,258	4,531	1,949	2,676	1,944
Per cent common equity	86.5%	75.3%	76.7%	79.7%	38.0%
Net profit	2,208	999	321	685	766
Dividends	1,051	648	184	330	517
Payout ratio	47.6%	64.9%	57.3%	48.2%	67.5%

* Includes long-term bank loans, pension reserves, and minority interests.

Source *Federal Reserve Bulletin* (June, 1956), pp. 586-87.

IMPLICATIONS FOR MANAGERIAL DECISION MAKING

In the course of these two chapters we have covered a great deal of ground and encountered problems still awaiting a solution which will satisfy the pragmatic business manager. Our approach has been to present the theoretical tenets underlying the problems involved and on which a suitable solution must be based, and to indicate, wherever possible, a practical method of treating these problems in the business firm. We now turn to the matter of defining or proposing an over-all policy approach to the capital planning function—an approach that seems logically to be called for in light of what has thus far been said.

Broadly speaking, the implications for managerial decision making can be summed up in a single sentence: *Management should pursue policies which are designed to reduce, or maintain at as low a level as possible, the company's cost of capital.* We will complete this chapter with a discussion of these policies.

Stockholder Relationships

The implications in this area are very broad, but can be discussed under two headings: (1) taking the stockholders into management's confidence, and (2) managing the corporation for the stockholders' benefit.

Taking the Stockholders into Management's Confidence. There is an old saying that "nothing succeeds like success," and, certainly in the field of business management, this old saw is as true as it might be anywhere else. A company that shows a consistently superior performance can, with respect to stockholder-management relations, pursue virtually any policy it wishes without doing any particular harm to the company's position in the capital market. An exceptionally good return earned on the stockholder's investment, achieved year in and year out, will induce most, if not all, stockholders to place complete faith in their company's management without any concern as to whether management tells them of the company's plans and objectives, or even whether the published financial statements are at all adequate for intelligent and critical analysis by the outsider. But such cases are exceptionally rare. Most companies cannot qualify in this regard, and so should be concerned with the state of existing relations, and should strive to improve it if there is a possibility of doing so.

A stockholder body which is kept well informed of the company's activities and objectives, which is treated explicitly by the management as the respected owners to whom the management is fully accountable for its acts, and which is given the clear impression that it is for the group's benefit that the corporation is being managed, is an extremely valuable asset. These stockholders are potentially the prime and most fruitful source of future capital, particularly equity capital (the expensive kind), and the favorable regard in which they hold their company and their stock can mean substantial dollar savings when outside capital is sought. It could even mean the difference between whether the funds will or will not be available in sufficient quantity and at a cost that makes the proposed capital projects worth undertaking.

Managing the Corporation for the Stockholders' Benefit. There is no doubt that the publicly held corporation has obligations to groups other than its owners. The community, the creditors, the management, the employees—all are vitally interested in the corporation's welfare and well-being. But only if corporate action is governed by considerations of stockholder interest will the correct economic decisions be made.[11] Thus, in considering a specific project proposal, we have seen that only if the project is evaluated in terms of whether it will increase the earnings available to the shareholders will the management make the correct economic choice.

In a broad sense, all managerial decisions can be related to this general heading of "stockholders' benefit." However, we may single out certain areas of managerial action that deserve special emphasis and discuss them as follows.

Dividend Policy

The most explicit and meaningful expression of the stockholder's relationship with his company is the dividend he receives on his stock. If only for this reason, then, the matter of dividend policy deserves the most thoughtful attention of top management.

In electing the corporation's directors, the stockholders delegate to them full authority over dividend policy, and it becomes a matter for the directors' discretion to determine what portion, if any, of the earnings shall be distributed to the owners. Recourse to the courts is always open to the shareholders, but this is a practical course of action only when there has been what amounts to an almost flagrant abuse of fiduciary responsibility. Even where earnings are very much larger than the current dividend, the penalty tax on improper surplus accumulation (imposed by Section 102 of the Internal Revenue Code) cannot be made to take effect as long as the earnings are put into physical plant, equipment, inventories, etc., or used to repay debt. Of course, as has already been pointed out, the consistently successful company with an outstanding performance record will not make its stockholders unhappy even with a very low payout policy, but we are not concerned here with the exceptional case.[12]

Because of the great importance which the typical stockholder attaches to his dividends, it is mandatory that a corporate management striving to improve its capital costs give consideration to the establishment of a dividend policy which will contribute to that end. The policy established should be consistent with the company's potentials and prospects. While dynamic growth is still going on, even a very low payout is justifiable, but an indiscriminate plowing of earnings into bank balances and government bonds is both unfair to the stockholders and harmful to the long-run position of the corporation. Briefly, for that is all that space permits, management should announce a policy which it intends to pursue and will stick to unless required by circumstances to deviate from it. This policy should include a decision to maintain a regular quarterly dividend which the company feels it can reasonably do, and an aim at improving the dividend whenever circumstances permit.[13] Many companies already pursue such a policy; too many others do not.

---

[11] In this area we begin to touch upon complex problems of public policy and regulation, but these controversial issues will be skirted because they are well beyond the scope of the matters that concern us in this chapter.

[12] Some readers might be quick to point out that the capital-gains-minded investor prefers price appreciation to dividend income. While this is certainly true, it is also a fact that stocks purchased primarily for growth suffer sharp price declines at the announcement of a dividend cut.

Just as it may be said that nothing succeeds like success, so we must also point out that in the business world there is no permanent substitute for success. A company can hardly make up for consistent losses by tackling the problem of dividend policy. Mediocre earnings will not suddenly blossom into large returns because regular dividends are paid out. But just as operating economies can be effected in production, materials handling, distribution, and marketing, so capital cost economies can be produced by giving adequate attention to such matters as dividend policy, and others discussed below.

Retained Earnings Policy

The decision to pay out 30 per cent of earnings is, at the same time, a decision to retain 70 per cent. Yet this area is important enough to justify a discussion of both aspects, while at the same time keeping in mind the fact that one cannot be decided without in effect deciding the other. The emphasis on dividend policy is intended to point up the need for management to recognize its responsibility to the corporate owners, and the desirability of developing a loyal stockholder body and a regard for the corporation which will redound to the long-run benefit of all concerned. The emphasis on retained earnings is intended to point up the importance of this major source of equity capital and its advantages over alternative sources. Because of the "double tax" on distributed corporate income, retained earnings have a lower net cost than do dividends returned to the company by way of new stock financing. But where the latter is a voluntary subscription, the retention of earnings amounts, in effect, to an involuntary subscription on the part of the stockholders. For this reason, retention policies should be carefully weighed against the desirability and benefits (to the corporation's long-run position) of greater dividend distributions.

---

[13] It does not follow that a company need adopt a ridiculously low regular dividend which it can feel will be payable out of earnings even in very poor years. Carrying this to its logical conclusion, it would appear that a company with a cyclical pattern of earnings, and which anticipates occasional losses, should pay no regular dividend at all. Some unlisted companies follow much that sort of policy, even though earnings would permit greater regularity. For example, in 1957 Lyon Metal Products, an important producer of industrial shelving and school lockers, paid a 15 cent quarterly dividend and a $3.40 year-end extra. On the other hand, Swift & Co., despite the cyclical nature of the meat-packing industry, embarked in 1950 on the ambitious policy of declaring a full year's regular dividend ($0.40 quarterly in that year, subsequently raised to $0.50 quarterly) in advance, supplementing it with special dividend payments at the end of the year. The company was forced to abandon this policy in 1958 after a very bad year in 1957, when the entire industry struggled to keep from going into the red. Still, the company's directors deserve commendation for a thoughtful approach to this important problem, and while the particular solution chosen is not suitable for a company susceptible to such wide swings in earnings, it might be a reasonable policy for more stable companies to adopt.

Table 11–1 showed the rising trend of equity costs, from which it obviously follows that the huge amount of internal financing that has taken place (referred to earlier in the section on Capital Structures) has been accomplished at almost a consistent increase in the cost of equity financing. But the personal income tax, as explained earlier in this chapter, places retained earnings at a great advantage over externally derived equity funds, and this is one of the basic principles on which retained earnings policy (dividend policy) should be built. The high cost of equity funds makes it mandatory that the project to be financed be rather highly profitable. Where retained earnings are employed, project profitability can be substantially less, and it would still be in the stockholders' interest that these projects be carried through. However, this should not be an excuse for drastic reductions in dividends and continued low payouts in all cases, unless economically justified. And even where such action seems desirable, management should advise the stockholders of the reasons for the dividend retrenchment and should give all necessary assurances that dividend improvement will be forthcoming as soon as possible.

The Corporate Income Tax

In attempting to meet head-on the problem posed by the corporate income tax, those firms contemplating the employment of new equity capital would undoubtedly find it desirable to be able to predict the future course of such taxes. Depending on one's analysis of the major forces involved, this particular prediction can be either very simple or very difficult. The former belief would be held by those who felt that corporate income taxes will remain high indefinitely, if not necessarily at the 50–52 per cent level; the latter opinion would be held by those convinced that conditions are always altering sufficiently to permit wide swings in tax rates from time to time. We are inclined toward the former view.

The forces involved are economic and political, domestic and international, changing and ever present. Pressing for continued high corporate taxes are:

1. The cold war situation which calls for a high level of defense spending.
2. A full-employment policy which, together with the organized labor movement whose leaders find it necessary to prove their worth by pressing for regular wage increases for the members, must necessarily produce price inflation.[14]

---

[14] This point is discussed in greater detail in Chapter 4.

3. The productivity of this revenue source and the difficulty of replacing it with another that would be politically less unpopular.
4. The increasing size of the "hard core" portion of federal expenditures —interest payments, farm subsidies, social welfare, highways, and numerous other demands that tend to become part of a permanent program rather than a "one-shot" proposition.

Arrayed against these forces is the basic and powerful desire on the part of everyone—corporations, stockholders, nonstockholders, low- and high-income groups—to minimize the tax bite. But whenever economic conditions seem to warrant a reduction in taxes, even stockholders would prefer a cut in personal taxes to a reduction in corporate assessments, and, in terms of popularity, an increase in personal exemptions would have much greater political appeal than a cut in corporate income taxes. Thus, while it may be that corporate taxes will come in for an occasional reduction, it is hardly expected that a return to the pre–World War II rates of about 15 per cent will be forthcoming. A 38 or 40 per cent rate is certainly a near-term possibility, but these rates are still high and will have great impact on the pre-tax cost of equity capital.

If the above conclusion is acceptable, it would follow that managements which have been traditionally conservative in their financing plans would do well to consider a shift in their approach and permit greater use of debt financing. A sharp shift in that direction is not necessarily called for, but where earnings typically indicate a substantial coverage of fixed charges (actual or potential) debt financing would be a reasonable course to pursue, though the traditionally conservative management might prefer to do so cautiously.

The Investment Timing Problem

An inseparable part of any investment decision is the timing of the expenditures involved. The wide swings that have occurred in the costs of capital are apparent enough from the data of Table 11–1, and the importance of timing can be more clearly brought home by a simple illustration. In September, 1957, Public Service Electric & Gas Co. (New Jersey) offered $60 million of thirty-year, first-mortgage bonds, rated Aa,[15] at a yield of 4.81 per cent. The bonds sold, but rather sluggishly. In March, 1958, only six months later, Union Electric Co. (Missouri) sold a thirty-year first mortgage issue, also rated Aa, to yield 4.22 per cent, and it was quickly taken by the market. This difference of 59 basis points (slightly over one-half of one per cent) meant that, had the Public Service company been able to market its bonds six months later, it would have saved $354,000 per year in interest charges—hardly a saving to be ignored.

In a broader sense, the timing problem is only one part of the bigger problem of forecasting. As related to capital expenditure planning specifically, it has already been shown that the investment decision is based on cash flow estimates (forecasts) expected over the life of the project. There arise, then, two considerations with respect to investment timing:

---

[15] Second only to the highest rating of Aa.

1. Planning the expenditure so that the new capital will be sought at opportune (low-cost) moments.
2. Selecting investments so that the cash flows are available for reinvestment at the most attractive rates.

Both of these aspects impose a demand on the decision maker to predict the future course and pattern of rates, and it would certainly be desirable to develop the skill and foresight that this would require. Suffice it to say, in this connection, that this is a goal worth striving for, but how difficult it is to attain has already been indicated in the above illustration, and may be further emphasized by the following data. On February 24, 1958, the Treasury's bill offering was priced to yield 1.202 per cent. This compared with a yield of 1.731 per cent just one week earlier, a 24-year high of 3.660 per cent on October 14, 1957, and a new low since that of 1.130 per cent on February 14, 1955. Putting this information in chronological order and computing the percentage fluctuations that have taken place in this issue (Table 11–5), we can readily comprehend the hazards involved in making the prediction of such trends a basic building block on which to rest capital expenditure decision making.

TABLE 11-5

FLUCTUATIONS IN SELECTED TREASURY BILL YIELDS

Date	Yield	Arithmetic Change	Percentage Change
Feb. 14, 1955	1.130%		
Oct. 14, 1957	3.660	+2.530%	+223.9%
Feb. 17, 1958.	1.731	−1.929	−52.7
Feb. 24, 1958	1.202	−0.529	−30.6

These fluctuations speak clearly enough for themselves, and emphasize the difficulties already indicated for those seriously hoping to time their capital procurement program scientifically. Let us turn now to the second aspect of the timing problem, before reaching any specific conclusions to serve as guides for managerial decision making.

The problem of selecting among investments with full consideration to the timing and availability of the generated cash flows is usually referred to as "the reinvestment problem." The point is made that it is not enough to determine that Project A provides a 15 per cent rate of return as against 12 per cent for Project B and that therefore A is to be preferred to B. For, the argument goes, it might be that the cash flow patterns are such that the heavier flows from Project A are anticipated at a time in the future when reinvestment opportunities are less attractive than those expected to exist at the time when the cash flows from Project B are made available. To take a simplified example, assume that both A and B have an economic life of two years, that A will produce one tenth of the expected cash at the end of the first year and nine tenths at the end of the second year, while B is expected to generate nine tenths of the total cash stream at the end of the first year and one tenth at the end of the second. Assuming that the size of the respective cash incomes were such that A indicated a higher rate of return, we would presumably select it rather than Project B.

But suppose the forecast indicates that reinvestment opportunities will be very attractive one year from today, and rather unattractive two years from today. This means, then, that a very large amount of cash will be available for attractive reinvestment if we select Project B, whereas most of the cash flow forthcoming from Project A will be available for reinvestment at substantially lower rates of return. Hence, the combined results of initial investment plus reinvestment of cash proceeds points to Project B as the more profitable alternative.

The argument as it stands is completely valid and cannot as such be refuted. But consider what this line of reasoning leads to. The corporation has a life in perpetuity so that the investment function it performs is a permanent and continuing one. Therefore, where is the justification in considering only a first stage of reinvestment of cash proceeds? Theoretically, all current investment decisions must be based not on a selection of profit alternatives as measured by rates of return on the respective investments, or even on these rates adjusted for the reinvestment returns on the cash proceeds as they become available in the future, but on a total investment and reinvestment plan that stretches indefinitely into perpetuity. Each investment produces cash flows which are, typically, reinvested and which in turn will produce other cash flows to be reinvested, and so on. This is what we mean by the permanent and continuing investment function of the corporation. Ideally, then, the optimum investment decision is that which selects from among an infinite number of infinitely long-term investment plans, that one which will produce the best investment results (verifiable only many many years later with hindsight knowledge).

There may exist, then, a discrepancy between the investments to be selected on the basis of the anticipated *rates of return* on the individual alternative projects, and the selection of alternative investment plans on the basis of the anticipated best *investment results* from a long-run continuing investment and reinvestment process. While the latter basis is the proper one upon which to build the investment function, the seeming discrepancy between the two alternative approaches may frequently be more imagined than real. This point is further amplified below.

Some Subtleties in the Rate-of-Return Analysis

Implicit in the rate-of-return analysis is the assumption that *each investment can be replaced at the end of its economic life by another of* *equivalent risk yielding the same rate of return.* This is a necessary premise in the analysis. Otherwise it is possible to make any investment superior to all others by assuming a reinvestment rate which will make it so. We will illustrate this implicit and basic principle in two important applications:

1. The borrower of $100 for ten years at 5 per cent compound interest might arrange to repay the loan in one of many ways. Two obviously equivalent alternatives would be: (*a*) have the interest added to the principal at the end of each year and make a single lump-sum payment at the end of ten years, amounting to $179.08; or (*b*) pay $6.00 annual interest charges, and repay the principal sum of $100 at the end of ten years. Both of these alternatives involve *compound* interest (though most people are likely to think that the latter case involves only simple interest), and what makes them exactly equivalent is the *implicit assumption* that the annual interest payments, as they are received in the second case, can be *reinvested at exactly the same degree of risk at 6 per cent.*

2. A bridge is to be constructed across a river and the choice (to simplify the illustration) is between steel and wood. A steel bridge would last, say, forty years; a wooden bridge would have to be replaced in, let us say, eight years. The advantage of the steel bridge is its lower annual maintenance costs. Its major disadvantage is the much larger investment required. Assuming that there is no preference for either type of bridge in terms of the quality of service, the investment decision will hinge on a comparison of annual costs (comprising maintenance costs and the investment's capital recovery cost), and the alternative involving the least annual cost would be the proper economic choice. The fact that the steel bridge would be much more durable does not alter the decision,[16] for implicit in the analysis is the assumption that the wooden bridge can be replaced after each eight-year period at *exactly the same* cost, maintenance disbursements, durability, and salvage value (if any), and that the same will be true for the steel bridge when the time comes to replace it. Thus, the annual costs computed for the original bridge is implicitly assumed to repeat indefinitely into the future.

This is, in short, an "other things being equal" approach. Where it is not possible to plan otherwise, or to predict whether the "other" things will change in one direction rather than another, it is the only meaningful approach to employ. In terms of the above example, the choice between the less durable wooden bridge and the longer-lived steel bridge is determined by what we consider today to be the correct economic decision. Our hope is that it will prove to be the right decision in the long run, and whether it does so prove to be will depend on economic and technological conditions eight years from now. Any factors that would make the replacement economics more favorable eight years from today (lower prices and wage costs, more efficient construction methods, more durable materials which cost as much or less than the wood does today) would be added reasons for selecting the wooden bridge today. Inflationary factors will make the choice of the steel bridge more attractive.

---

[16] Another implicit assumption in this problem is that the services of a bridge will be required indefinitely. If we impose, however, the assumption that, for whatever reason, a bridge is needed for only fifteen years, this will have to be considered in making the correct choice.

Thus, "the reinvestment problem" as stated earlier is not a newly discovered one, and is only one aspect of all investment decisions which must be made today in a continuing and perpetuating society or firm. The sophisticated decision maker takes all factors into consideration, whenever he can, and acts accordingly. The more dynamic the industry and the more uncertain the external forces which act upon the firm, the greater the number and importance of the factors likely to cause discrepancies between anticipated and actual results. In such cases, investment decisions are likely to be biased in favor of very short payouts. In more stable and predictable industries (such as public utilities), long payout investments are much more common.

Notwithstanding the theoretical weight of the foregoing discussion, there is something to be said for limiting investment decisions to a strict rate-of-return analysis without encumbering (and endangering) the analytical process with what are frequently rather tenuous and intractable factors.[17] This is not a denial of the desirability of estimating these factors whenever possible; it is rather an expression of caution against attempting to put too much weight on these estimates in the typical investment decision. Certain investment processes, e.g., mutual investment trusts, are more suited for this type of decision making; it is not so for the typical manufacturing enterprise. In the latter case, there is, in the first place, substantially less fluidity in the asset structure, which in turn places greater emphasis on the need for making the correct decision *today* rather than being able to correct today's wrong decision tomorrow (the "sunk" nature of capital investments); and, in the second place, the investment results of a business enterprise over time are much more importantly dependent on the correct estimate of the cash flows expected from a given project than upon the reinvestment of these cash flows at opportune moments "brilliantly" anticipated (more likely by pure chance, if at all).

To state the last thought somewhat differently: those who point to the "reinvestment problem" as a discrepancy-producing source in the rate-of-return analysis are either reflecting an excessively pessimistic view of the availability of attractive investment opportunities, or are with gross optimism suggesting that the difficult problem of estimating profitability can be further compounded by adjusting the rates of return by the profitability of reinvesting cash proceeds of varying pattern over a variety of time spans, without at the same time making the problem an impossible one. For the pessimistic view it need only be said that there is never, at any time, a shortage of profitable investment opportunities—there is more likely to be a shortage of risk takers capable of seeing the opportunities that are available; and while a given investment opportunity might more profitably (opportunely) be undertaken at one time than at another, there are always other opportunities whose right moment is *now.*[18] For the optimistic view we will repeat, in somewhat different terms, the point made earlier, *viz.*, why stop at the "first stage" of reinvestment? If it will improve investment results to modify the dictates of the rate-of-return analysis by what the reinvestment estimates indicate, why not carry the logic still further with a third, fourth, and fifth modification of what the reinvestment of the reinvested proceeds (etc.) would indicate? Obviously, it is all a matter of degree of difficulty. But our point is that stopping at the "first degree"—with the rate-of-return analysis—is as far as we should *ordinarily* go. In an investment process where it is possible to go beyond this, it should be done. Otherwise, the intrepid forecaster who is willing to venture into the dark unknown should be wise enough to place a heavy premium on the rate-of-return results, and give only slight weight to the suggestions of the reinvestment estimates. In a sense, this tends to be done in the business decisions biased in favor of short payouts.

---

[17] Considerations such as those indicated in the bridge illustration above, i.e., inflation factors versus technological improvement, are relatively so simple that they can be, and usually are, readily taken into account in any investment decision. We have in mind at this point much more difficult considerations relating to the desirability of a particular pattern of cash proceeds and the relative profitability of one pattern as against a number of others because of the reinvestment opportunities envisioned at different points in the future.

Public Relations with the Financial Community

We have already discussed under a separate heading the matter of stockholder relationships, wherein was emphasized management's responsibility to the owners. In this section we concern ourselves with the corporation's financial public relations with the general financial community in which investment opinion is formed and develops and which, in turn, can have tremendous significance to the future of a business enterprise.

---

[18] We referred earlier to the recent postponements in 1957 of capital expenditure plans by General Electric, General Motors, and Aluminium Ltd., and of the "stretch-outs" by Alcoa and Reynolds Metals. These alterations in capital plans do not reflect, however, a managerial decision to take advantage of more opportune reinvestment of future cash proceeds. They are, rather, an attempt to cope with near-term uncertainty as it affects the decision to carry through a specific project. It is important that the reader note the essential difference involved. Temporarily shelving a given project because of near-term uncertainty for a later time when the outlook is a little clearer is far different from trying to select from among a set of alternatives that project which will produce a pattern of cash flows most opportunely available for a series of future reinvestment.

Departing from Tradition. Product advertising—employing aggressive, persistent, and repetitive techniques—is a widely accepted part of promoting business and expanding profits. Yet, many of the companies that are well known for "hard selling" of their product lines have been manifestly unaware (at least they have not done anything to show their awareness) of another type of selling—cultivating the investor!

The background and traditions of business practices and attitudes probably explain this difference. To make profits it is necessary to sell the company's products and services and, up to a point, the more that can be sold at a given price per unit the greater will be the profit. Since products and services do not sell themselves, successful selling techniques become an important part of this logic. On the other hand, competition has traditionally induced businessmen to feel that the less known by those outside the management the better, with the result that a "none-of-your-business" attitude toward the outside investors has tended to prevail, though it has been gradually giving way to a more enlightened attitude of telling investors all that possibly can be told.

Advantages of a Broad Ownership Base. The spread of public ownership of most large corporations has brought an awareness to many managements that a change in attitude was necessary. Among the first companies to recognize this change and take advantage of it were those engaged in selling products directly to consumers (foods, oils, autos, appliances, beverages, etc.). Realizing the potential market represented by the stockholders, many of these companies have made it a point to advertise their products in the annual reports and other materials which are sent out to the owners, and to maintain, by means of stock dividends and splits, as broad an ownership base as possible by keeping the price of the stock within a reasonable buying range. High-priced stocks are unpopular with investors; stock splits bring the price down to a range that will have broader appeal. Such considerations are important: the purchaser of an automobile, for example, will, other things being more or less equal, give preference to the product of the company in which he is a stockholder. There are other advantages to a broadened ownership base: (1) it facilitates retention of control by making it more difficult for ownership to become concentrated in unfriendly hands; (2) it means a larger source of equity and debt capital, or at least a larger ready market to which an appeal for such capital can be made; and (3) it provides added incentive to independent financial services to give some attention to the affairs of the company.

Required—A New Selling Technique. The financial public relations program we recommend involves the adoption of a long-run policy to which the firm must commit itself unwaveringly, and cannot be switched on and off as readily as one switches advertising media and techniques. The approach required is radically different from that of product advertising:

1. The items for sale are the company's financial record and its future prospects.
2. The "market" to which the appeal must be directed is made up of investors, the most important segment of which is the institutional investor. The latter is an "enlightened professional," so to speak, with large resources for investigating and analyzing situations, and many connections for checking information.
3. The underlying philosophy must be a willingness to deal as honestly and openly as possible with the present and potential investors in the company's future.

Rules of Conduct. It is impossible, of course, to consider here in detail the many ramifications of this subject. The arguments are not all on one side, but there seems to be taking place an awakening, so to speak, on the part of American managements to the importance of cultivating the investor. This is at least evidence of the fact that an increasing number of managements have come to the conclusion that the effort might prove rewarding. It is no simple matter to lay out a program that will be suitable for all companies alike, but some generalities are possible. The basic approach is what might be described as "sincere aggressiveness" on the part of the management to secure for the company the advantages of a favorably inclined capital market.

1. Hold regional stockholder meetings (depending on the size of the company and the concentration of stockholders in various parts of the country). This gives many of the stockholders an opportunity to see their management, hear their plans (in a general way, of course), and ask questions. Contrast such an approach with the current policy of so many large corporations which select a rather inaccessible, out-of-the-way meeting place for the annual stockholder meeting—obviously an attempt to minimize stockholder attendance.

2. Aggressively seek, and graciously accept, invitations to speak at the various investment analysts societies throughout the country. The members of such societies are typically the representatives of the largest institutional investors in the country—major banks, insurance companies, and mutual funds—as well as representatives of brokerage firms and investment counsellors whose reactions ultimately are made known to most investors active in the market. At these meetings, executives are invited to discuss their company's operations, plans, problems, and prospects. These forums present an ideal opportunity for a profitable selling job of the executive talent, and the company's future. Social contact of this sort between the company executives and the investment analysts adds an element of realism to the performance of the latter's duties and makes the company

more than just a "name." From the company's point of view, a presentation effectively carried out can enhance its standing in the investment community, though any permanent enhancement would have to be supported by operating results over the future.[19]

3. Build investor confidence by requiring all management personnel to own and retain a reasonable minimum amount of stock in the company, such stock to have been acquired by direct purchase in the market over a period of five years (so as not to force acquisition at times when stock prices are generally believed to be too high). This ownership should be related to the individual's position and salary, *and should come outside of any stock option plan that the company might have.* The latter is a justifiable means, when reasonably employed, of compensating management for a job well done, but is not an expression of management's confidence in the business. A properly expressed vote of confidence by management will go a long way to calling forth a vote of confidence by the stockholders in the company and its management (in the form of higher price-earnings ratios on the stock, and a ready willingness to subscribe to new security issues at prices attractive to the company).

4. In meeting with a group of investors, whether actual or potential, management should discuss their questions frankly and approach them with a ready willingness to state some of the problems confronting the company—how long it seems likely it will take for the solution of problems to be forthcoming, the amount of sales and earnings currently being budgeted for, a clear-cut statement of the company's present and probable dividend policy, and the direction in which the company expects to move in the future. The investors are actually entitled to such treatment, and will react favorably toward a management which accords it to them. This will show itself in enhanced stockholder loyalty and a generally better price-earnings ratio—important considerations, of course, when new capital is sought.[20]

The Need to Sell Success. We made the point earlier in this chapter that nothing succeeds like success. This is generally true, but we must modify this adage by the comment that sometimes success has to be sold. The psychology of the investor—even the well-staffed institutional investor—can cause him to act in peculiar fashion. How else explain the fact that company A has a much better record of dividend and earnings growth than company B in the same or allied industry, yet consistently sells at a significantly lower price-earnings ratio? To a great extent this can be due to a lack of close familiarity with the undervalued company on the part of the investment community. No investor, institutional or otherwise, is able to follow very closely the affairs of all companies listed on the securities exchanges, let alone those of the many unlisted companies. The result is that a given company might be undervalued indefinitely, in spite of a consistently good operating record, and the only remedy for such a situation is to "sell" the company to the investment community by means of a wisely conceived financial public relations program comprising most of the suggestions put forth herein.

---

[19] In February, 1958, Mr. R. E. Reimer, Executive Vice President, Dresser Industries, appeared before the St. Louis and Chicago societies. His frankness in discussing the company's problems, his obvious enthusiasm for his company's long-term outlook, and his willingness to state so early in the year his expectations of the company's earnings in 1958 was indeed refreshing, and is the sort of approach which other publicly held companies might seriously study and consider.

[20] Companies that attempt expansion via the merger route would also benefit handsomely by a high price-earnings ratio on their stock. While many factors may determine the exchange ratios in a merger, current market prices are probably the most important single determinant. Therefore, the merging company, with stock selling at a low P/E ratio, will have to pay out that much more stock to acquire another firm, thus resulting in dilution of earnings and book value.

CONCLUSION

If there is one underlying theme to this book, it is that planning for the future is an essential ingredient in the successful long-run operations of a business enterprise, and that the type of planning most likely to produce the desired long-term results is that which is based on the recognition that production today for the uncertainty of tomorrow requires an objective approach to the problem: the making of concrete estimates of markets, costs and profits. Notwithstanding the emphasis placed on the policy implications and suggestions discussed above, each of them, or in combination, is a very poor substitute for good operating results, and can serve a useful function only as a supplement to management's efforts to improve the company's operations.

The details of planning procedures, their administrative aspects, the accounting techniques involved—these are all beyond the scope of this book. As for the actual planning and forecasting methods themselves, there are several from which to choose. Most firms can avail themselves of the simplest techniques; some have already adopted the most sophisticated ones. The latter firms are, so to speak, working on the frontiers of forecasting and planning—an expense which only the rather large corporations can afford.

Our aim has been to correlate the traditional theoretical concepts of the economics of the firm with the planning and forecasting techniques available for the solution of some of the specific business problems encountered by an operating company. We have presented actual studies which showed how these techniques have been used for the solution of some business problems, and have tried, where possible, to offer a simple, short-cut approach as well, of which the smaller, not-too-well-financed, firm could avail itself. Although we have avoided reference to more advanced (mathematical) techniques, to the hardheaded (practical) businessman, and even to many students who will read this book, some of the techniques which we did present will probably seem a little like reaching for the moon. But then, "a man's reach *should* exceed his grasp or what's a heaven for"?

BIBLIOGRAPHICAL NOTE

Capital theory, which forms part of a broader treatment of economic theory, is nevertheless the basis of modern capital budgeting. The historical development of economic thinking on the subject may be surveyed briefly with the following works: Eugen von Böhm-Bawerk, *Positive Theory of Capital;* K. Wicksell, *Lectures on Political Economy;* I. Fisher, *The Theory of Interest;* F. Knight, "Capital, Time, and the Interest Rate," *Economica* (August, 1934); K. E. Boulding, "The Theory of a Single Investment," *Quarterly Journal of Economics* (May, 1935), and his *Economic Analysis,* 3d. ed., chap. 39; J. M. Keynes, *The General Theory of Employment, Interest and Money;* P. Samuelson, "Some Aspects of the Pure Theory of Capital," *Quarterly Journal of Economics* (May, 1937); J. R. Hicks, *Value and Capital,* 2d. ed., and F. and V. Lutz, *The Theory of Investment of the Firm.*

In addition to these theoretical works, a number of studies of capital expenditures planning have been made in recent years and reported in various sources. Those employing the interview technique, in addition to the works cited earlier in Chapter 2, include: R. P. Mack, *The Flow of Business Funds and Consumer Purchasing Power;* W. Heller, "The Anatomy of Investment Decisions," *Harvard Business Review* (March, 1951); M. Gort, "The Planning of Investment: A Study of Capital Budgeting in the Electric Power Industry," *Journal of Business* (April, 1951); and R. Eisner, *Determinants of Capital Expenditures,* University of Illinois Bulletin, 1956. An empirical analysis of investment decision making along with its theoretical implications is presented in an excellent recent work by J. R. Meyer and E. Kuh, *The Investment Decision.*

For the most abundant recent contributions relating directly to the subject matter of these two chapters, the *Journal of Business* and the *Harvard Business Review* are the richest sources. The October, 1955, issue of the former publication is devoted entirely to capital budgeting, with the contributions by Ezra Solomon and M. J. Gordon being particularly of interest. Other *Journal of Business* articles worthy of mention include: E. Solomon, "The Arithmetic of Capital-Budgeting Decisions" (April, 1956); H. V. Roberts, "Current Problems in the Economics of Capital Budgeting" (January, 1957), and, in the same issue, G. Shillinglaw, "Profit Analysis for Abandonment Decisions"; E. Renshaw, "A Note on the Arithmetic of Capital Budgeting and the Problem of Reinvesting Cash Proceeds" (October, 1957). Recent articles of particular note in the *Harvard Business Review* include: D. B. Woodward, "Regularizing Business Investment" (May, 1952); R. P. Soule, "Trends in the Cost of Capital" (March, 1953); J. L. Peirce, "The Budget Comes of Age" (May, 1954) E. G. Bennion, "Capital Budgeting and Game Theory" (November, 1956); and R. Reul, "Profitability Index for Investments" (July, 1957). An earlier treatment of capital budgeting, as already mentioned in the previous chapter, was Dean's little book on the subject, the essence of which appears in his *Managerial Economics,* chap. 10.

QUESTIONS

1. Why is the supply and demand technique, as typically employed in equilibrium price analysis, not suitable for determining the appropriate capital expenditure quantity of a firm?
2. What is a firm's potential internal supply of capital?
3. Assuming a firm with no long-term debt or preferred stock, draw a curve which depicts the availability of debt capital to that firm, and explain its meaning. State your assumptions. If this firm were to acquire more equity, what effect would this have on the curve you have just drawn?
4. The L Company follows a policy of paying 35 per cent of cash earnings in dividends, and reinvesting 65 per cent in replacement of existing facilities and for expansion. Assuming that this is their guiding principle in determining capital expenditure planning, what name is given to this method of resource allocation? What do you think of this method? Can you see why, if used in a more sophisticated manner along with other guides, it might be a very suitable (practical) approach to the problem of capital budgeting?
5. Discuss the tendency toward "debt aversion," and relate this to the premise of profit maximization. What recent developments have lessened, somewhat, management's aversion to debt?
6. Distinguish between the "borrowing rate" and the "lending rate," as put forth by F. and V. Lutz. Are these rates ever the same? Are either or both, opportunity cost rates? What have they, if anything, to do with the firm's cost of capital?
7. The economic literature makes frequent use of the term "the interest rate." Is there such a thing? If so, what is it? If not, what then does it mean? In the same sense, what is "the borrowing rate" and the "lending rate" mentioned in the above question?
8. Do retained earnings involve a cost to the firm employing them? Why? Does the cost, if any, depend at all on whether dividends are being paid? Why or how?
9. Contrast, in a given firm, the costs of alternative sources of capital. What assumptions are necessary for making such a statement?
10. Distinguish between "earnings yield," as that phrase is typically employed, and rate of return on investment.
11. "A truly democratic policy which all businesses should pursue is 100 per cent distribution of cash earnings, the corporate capital being replenished by offer of stock subscriptions. Only in this way would the stockholders truly control the corporation's capital, and the free market forces will truly allocate resources among competing firms in the most optimum manner." Discuss.
12. What is meant by trading on equity? Why is this principle, as it is usually stated, an inadequate guide for appropriate managerial action for maximizing profits?
13. Explain what is meant by "the marginal real cost of borrowing." How is this concept useful in defining optimum capital structure? What is important about capital structure as it relates to a firm's cost of capital?

14. Describe the capital cost patterns since 1920. What are the major factors in the costs of debt capital and equity capital respectively?
15. You have been appointed to the post of director of public relations for your firm, a publicly held corporation. The president has asked you to study and present in writing what you feel would be an integrated public relations program aimed at improving the firm's cost of capital and accessibility to the capital markets. What are your recommendations?
16. What is meant by the "reinvestment problem"? Discuss in full, indicating the nature of the "problem," the proposed "solution," and the shortcomings of the "solution."

SUPPLEMENTARY PROBLEMS

The following pages contain a set of problems of a quantitative nature. These problems are of two types:

1. Those that are virtually identical to problems illustrated in the text.
2. Those that require what might be called "thought flexibility" in that they have not been specifically illustrated in the text but can be solved by planning a logical method of attack beginning with the underlying principles provided in the book.

This supplementary section may be considered optional by the instructor. The problems are generally more difficult than the questions which appear at the end of each chapter, but they are by no means extraordinarily difficult for it would be quite reasonable to expect the average student using this book to solve them all with some effort. However, they do take time to solve and can typically be expected to consume the better part of a class period for full discussion and analysis. In that sense they might be used as "quantitative case material," to be supplemented, if the instructor should wish to do so, with problems and cases of a qualitative nature available in such books as:

M. R. Colberg, W. C. Bradford, and R. M. Alt, *Business Economics* (Rev. ed.; Homewood, Ill.: Richard D. Irwin, Inc., 1957).

T. C. Raymond, *Problems in Business Administration* (New York: McGraw-Hill Book Co., 1955).

I. (Chapters 3 and 5)

The American Resort Hotel Association, a national organization of better resort hotels, is interested in providing for its members any information that would be of use in helping to determine the location of new hotels. Assume that you have obtained from the Association and from other sources shown in the footnotes the information presented in Table A.

1. What is meant by a "guest-week"?
2. Using graphic techniques, prepare a preliminary analysis of the relationships between the variables. Label your charts "A" and "B" and state below it what it is that the chart reveals. Use drift lines if necessary in guiding your estimate of the regression line.
3. Is this a cross-sectional or historical type of study? Can you plot the results as a time series? Explain.
4. Assume that this is a historical study. Then in column 1 we can substitute "years" instead of "hotels." Thus we would have "year 1" instead of "A," "year 2" instead of "B," and so forth. Do this, and then:

a) On a chart, plot the actual and calculated results between Y and X_1, using a solid and a dashed line, respectively. Label your graph Chart C. State below the chart what it reveals.

b) On another chart, Chart D, plot the actual and combined estimates, or the $Y - X_1, X_2$ relationship, again using a solid and dashed line. Explain below the chart what it reveals.

c) Is there a substantial difference between Charts C and D? Explain.

d) Compute the multiple coefficient of determination and interpret your result.

TABLE A

GUESTS PER SEASON, PER CAPITA INCOME, AND AVERAGE RATES PER PERSON FOR SELECTED RESORT HOTELS IN THE UNITED STATES, SINGLE SEASON

Hotel	Guest-Weeks for Summer Season	Per Capita Income*	Average Rates Per Person Per Week†
A	7,310	$1830	$150
B	10,901	1935	134
C	11,782	1950	120
D	9,625	1965	136
E	8,500	1992	142
F	9,309	2040	140
G	8,921	1785	134
H	5,673	1500	136
I	3,338	1140	142
J	3,485	1104	140
K	4,200	1254	130
L	8,723	1380	96
M	9,405	1593	110
N	9,500	1683	112
O	7,802	1530	116
P	6,291	1623	120
Q	5,490	1737	146
R	9,085	1976	140

* Within 150-mile radius of each hotel. Estimated from published data of *Sales Management* Magazine and U.S. Department of Commerce

† All rates are for American Plan (i e , meals included) and have been adjusted to reflect both single- and double-room rates per person.

e) Assuming a per capita income of $1,900 and an average guest rate of $140 per person per week, forecast on Charts C and D the expected guest-week figure for year 19.

II. (Chapter 4)

You have been appointed business economist for the Grandview Manufacturing Corporation.

1. On the basis of the following profit and loss data, construct a break-even chart showing the various break-even points. (Hint: Consult the bibliographical note at the end of Chapter 4 for further help, if necessary.)

Net sales	$1,000,000
Variable cost of sales	400,000
Fixed cost of sales	100,000
Gross profit	$ 500,000
Variable selling and administrative expense	100,000
Fixed selling and administrative expense	50,000
Net operating profit	$ 350,000
Bond interest	50,000
Net to stockholders	$ 300,000
Preferred dividend requirements	50,000
Net to common stockholders	$ 250,000
Regular common dividend	150,000
Profits to earned surplus	$ 100,000

2. What assumptions are involved in constructing a break-even chart?
3. How can the chart be used in forecasting?

III. (Chapter 4)

A. Construct a simplified example which illustrates the "smoothing out" effect on net income resulting from LIFO inventory accounting, and contrast this with the "exaggerated" effects of the FIFO method. Show what happens in a period of rising prices, and in a period of falling prices.

Suggestion:

Assume initial inventory of 100 units, sales of 100 units per year, and purchases of 100 units per year. Assume initial inventory is costed at 80 cents per unit, but cost rises to a level of 90 cents for the entire first year and closes at that price; then rises to $1.00 for all of the second year; and back to 80 cents for all of the third year. Selling price each year is assumed always to be 10 cents per unit higher than cost in that year. No other costs are involved. The "lower-of-cost-or-market" rule applies.

B. Construct a situation, illustrated arithmetically, under which LIFO accounting results in a distorted picture of operating profits, i.e., in a period when one would expect to see losses or very low profits at best, the LIFO method could, under the right circumstances, produce surprisingly good profits. The "lower-of-cost-or-market" rule applies.

Suggestion:

1. First see what happens if in the second year the cost per unit had fallen to 60 cents (instead of rising to $1.00).
2. Then, returning to the original data, assume that in the third year the cost rose to $1.20 per unit (instead of falling to 80 cents per unit); assume also that the company's plant is struck so that no purchases are made or processed, and that all sales (continuing at the same annual rate) are out of past inventory.

IV. (Chapters 4, 10, and 11)

If you paid $990 for a $1,000 bond maturing in exactly two years, on which interest of $40 is payable annually, the first of the two payments being due one year from today, what would be the yield or return on your investment? Would the yield be different if the same bond were maturing in ten years rather than two years from today? Why or why not? Suppose the same bond had no maturity date, how would you compute the return on investment in that case? State any assumptions you must make in answering the last part of this question.

V. (Chapter 6)

Refer back to question 3 at the end of Chapter 6, p. 231. If you have not yet done this question, do it now.

VI. (Chapters 6 and 7)

Given the following data, construct a cost schedule showing total fixed cost, total variable cost, total cost, average total cost, average variable cost, and marginal cost. Plot the results and label the curves. (Note: plot the total curves on one chart, and the average and marginal curves on another, since different scales will be needed.) Assume that the fixed investment amounts to $1,000 and that labor costs are $10 per unit. From your table, derive the equation for total cost as a function of labor input.

Labor input	0	1	2	3	4	5	6	7	8
Quantity of output (tons)	0	4	12	17	20	21	22	23	24

VII. (Chapter 7)

Joel Dean estimated the total cost function ($\$C$) for a hosiery mill as $C = -10{,}485 + 6.75Q - 0.0003Q^2$, where Q represents quantity produced.

1. Write the equation for average total cost, A.
2. Find the total cost at:
   a) $Q = 6{,}000$.
   b) $Q = 4{,}400$.
3. Find the average total cost at:
   a) $Q = 5{,}000$.
   b) $Q = 8{,}200$.
4. Plot the total cost curve and the average cost curve for outputs ranging from 1,000 to 10,000.

VIII. (Chapters 7 and 8)

An independent automobile manufacturer has a capacity of 300,000 cars per year. After a few months have passed, the management estimates that domestic sales, at $2,000 per car (f.o.b. the factory), will be in the neighborhood of only 100,000 cars. Assuming the cost and revenue data of Table B, the management is confronted with the advisability of selling incremental output in foreign (differentiated) markets at a price which will at least defray some of the loss on the domestic operation.

1. Fill in all the blank columns.
2. Construct an average revenue (demand) curve which assumes that an incremental output of 50,000 cars will be "dumped" at $1,500 per car.
3. On the same graph construct an average cost curve.
4. Assuming a strictly domestic operation, what is the "break-even" point?

TABLE B

COST AND REVENUE DATA
(In thousands)

OUTPUT	REVENUE	FACTORY COSTS		ADMINISTRATIVE AND SELLING COSTS	TOTAL COST	AVERAGE CAR COST
		Direct	Indirect			
0		$0	$145,000	$ 80,000		
25		17,000	148,000	85,000		
50		34,000	151,000	90,000		
75		51,000	154,000	95,000		
100		68,000	157,000	100,000		
125		85,000	160,000	105,000		
150		102,000	163,000	110,000		
175		119,000	166,000	115,000		
200		136,000	169,000	120,000		
225		153,000	172,000	125,000		
250		170,000	175,000	130,000		
275		187,000	178,000	135,000		
300		204,000	181,000	140,000		

5. Assuming again the possibility of "dumping," what is the lowest price which the company might reasonably set for the export markets (f.o.b. the factory) in order to defray the loss on the domestic sale of 100,000 cars?
6. At a dumping price of $1,500 per car, what is the loss incurred on the total operation? What is the break-even point, assuming a 100,000-car domestic market, with the excess to be sold abroad at $1,500 per car? How does this loss compare with what would be incurred if no dumping were engaged in?
7. What is the lowest dumping price which would permit a break-even operation, assuming still that only 100,000 cars are sold domestically?

IX. (Chapters 10 and 11)

An industrial property consisting of land, plant, and equipment, is purchased for $80,000, and is improved, before being leased, at an additional outlay of $20,000. The lease agreement calls for rentals of $12,000 per year for the first three years, $13,000 in the fourth year, and $14,000 in the fifth year. At the end of five years, the lease would be subject to renegotiation. During the lease period the lessor agrees to pay certain expenses involved, such as insurance and property taxes, and these come to $2,000 per year in the first two years, and $3,000 in each of the last three years.

At the end of the five-year term, the lessee decides not to renew the lease, and the lessor succeeds in selling the property through a broker at a price which nets him $150,000 after paying a commission of 5 per cent on the gross sale price.

Assume that the initial purchase and improvement outlays were made at the *beginning* of the five-year period, and that all other expenses and revenues were respectively incurred and received at the *end* of each of the years in question.

1. What was the gross sale price?
2. What was the rate of return on the investment?

   (Note: see the text for any data needed to solve this problem.)

X. (Chapters 4, 10, and 11)

A twenty-year loan of $10 million is arranged under the following terms:

1. No interest to be paid during the first five years.
2. Starting at the end of the sixth year, interest of $550,000 per year is paid each year through the end of the twentieth year.
3. The principal of $10 million is to be paid at the end of the twentieth year.

Making use of any of the information given below, calculate the correct rate of interest on this loan as accurately as the information allows.

	4½%	*3½%*	*5½%*	*4%*
Present value of $1.00 per year, for 5 years. . . .	$ 4.390	$ 4.515	$ 4.270	$ 4.452
Present value of $1.00 per year, for 15 years	10.740	11.517	10.038	11.118
Present value of $1.00 per year, for 20 years	13.008	14.212	11.950	13.590
Present value of $1.00 five years from today . .	0.803	0.842	0.765	0.822
Present value of $1.00 fifteen years from today	0.517	0.597	0.448	0.555
Present value of $1.00 twenty years from today	0.415	0.503	0.343	0.456

Indexes

AUTHOR INDEX

A

Adelman, M. A., 364
Arrow, K. J., 24

B

Backman, J., 320
Banks, S., 160
Baumol, W. J., 24
Beach, E. F., 82
Bennion, E. G., 48, 83
Berle, A. A., 132
Bishop, F., 276
Black, J. D., 82
Blair, M., 230
Bober, M., 230
Boehm-Bawerk, E., 437
Borden, N., 277
Boulding, K., 389, 437
Bratt, E., 48, 82
Bronfenbrenner, J., 48
Bross, I. D., 24, 48
Brunk, M., 82

C

Carlson, S., 230
Cassels, J., 230
Chamberlin, E. H., 276
Christ, C., 48
Clark, J. B., 88
Clark, J. M., 211, 233, 241–42, 276
Cobb, C., 212, 230
Colberg, M., 48, 133, 198, 231, 277

D

Dean, J., 48, 132, 231, 266, 276, 319, 377, 437
Derksen, J., 198
Devine, C., 132, 277
Dirlam, J., 365
Douglas, P., 212, 230

E

Earley, J., 319
Edwards, C., 365
Eisner, R., 437
Ellis, H., 82
Ezekiel, M., 82, 254, 256–60

F

Federer, W., 82
Fellner, W., 132
Ferber, R., 82
Ferguson, A., 276
Fisher, I., 437
Fox, K., 198
Friend, I., 48

G

Gardner, F., 132
Gaumnitz, R., 24
Gayer, A., 230
George, H., 89
Gordon, M., 437
Gordon, R., 132
Gort, M., 437
Guthmann, H. G., 133

H

Haley, B. F., 132
Hall, R., 319
Hansen, H., 231
Hart, A. G., 24
Hawkins, E., 319
Hayes, J., 276
Heady, E. O., 24
Heller, W., 437
Hicks, J. R., 437
Hitch, C., 319
Hollander, S., 198
Howard, J., 48, 133, 277

J

Jacoby, N., 132
Jones, N., 132

K

Kahn, A. E., 365
Kaldor, N., 230
Kaplan, A. D. H., 364
Katona, G., 48
Kennedy, R., 133
Keynes, J. M., 147–49, 389, 437
Klein, L., 48
Knight, F., 24, 437

L

Lester, R., 319
Lever, F., 276
Lewis, W., 319
Lorie, J. H., 48, 82
Lutz, F., 377, 403, 405, 437

M

Machlup, F., 319, 366
Mack, R., 437

Malenbaum, W., 82
Marschak, J., 198
Marshall, A., 398
Marx, K., 89, 94
Means, G. C., 132
Meyer, J., 48, 437
Miller, H., Jr., 48
Modigliani, F., 48
Moore, G., 33
Mosak, J., 276
Mund, V., 366

N

Newman, P., 230
Nickolls, W. H., 209–10, 213–18, 230
Noyes, C., 276
Nutter, G. W., 364

O

Owens, R., 24
Oxenfeldt, A., 320

P

Papendrou, A., 365
Paradiso, L., 198
Peirce, J., 437
Phelps, D., 231
Pigou, A. C., 320
Powlison, K., 132
Prest, H., 198

R

Renshaw, E., 437
Reul, R., 437
Rickey, B., 3
Roberts, H., 82, 437
Robinson, J., 320
Roos, C. F., 48, 198, 319

S

Samuelson, P., 437
Sauerlander, O. H., 48
Saxton, C., 319
Schultz, H., 198
Schumpeter, J., 89–90
Schweiger, I., 48
Shackle, G. L. S., 24
Shepherd, G., 82
Shillinglaw, G., 132, 437
Smith, H., 276
Solomon, E., 437
Soule, R., 437
Spencer, M., 230
Sprowls, R., 48, 82
Sraffa, P., 230
Staehle, H., 276
Stelzer, J., 366
Stigler, G., 198, 230, 276, 319, 365
Stocking, G., 366
Stone, R., 198
Sweeney, H. W., 108
Szeliski, V., 198

T

Tannenbaum, R., 24
Thomsen, F., 82
Tintner, G., 24, 48, 276

V

Veblen, T., 89
Vickrey, W., 144

W

Walker, F., 88
Watkins, M., 366
Weintraub, S., 24, 132, 230, 319
Welsch, G. A., 132
West, V. I., 170–72
Weston, J. F., 132
Wheeler, J., 365
Whitman, R., 191–92
Wicksell, K., 437
Wilcox, C., 365
Williams, O., 230
Woodward, D., 437
Working, F., 198
Wylie, K., 254

Y

Yntema, T. O., 254–60, 192–94

SUBJECT INDEX

A

A & P case, 353–55
A priori deduction, 5; *see also* Risk, objective prediction
Advertising, 154–60
  and budgeting, 268–73
Alloted funds, 400–402
Antitrust laws
  applied to combination and monopoly, 330–32
    distribution, 352–55
    patents, 332–37
    price discrimination
      delivered pricing, 347–52
      discounts, 341–47
      general legality, 339–41
    resale price maintenance, 355–59
    restrictive agreements, 329–30
    trade marks, 337–38
    tying contracts, 338–39
  enforcement of, 326–28
  provisions of, 323–26

B

Basing point system, 347–52
Borrowing rate, 403–4
Break-even analysis, 109–16

C

Capital
  cost of; *see* Cut-off rate
  internal supply of, 399–400
  sources of, 95
Capital budget
  approval process in, 370–72
  construction of, 370
  nature of, 368–70
Capital expenditure criteria
  annual cost, 383–86
    of equipment, 385–86
    in real estate loans, 383
  capital recovery period, 382–83
  payout, 381–82
    and corporate income tax, 382
  rate of return
    approximations to, 389–95
    defined, 387
    and marginal efficiency of capital, 389
    simplified version of, 387
  urgency or postponability, 380–81
Capital management
  administrative aspects of, 368–72
Capital management—*Cont.*
  and capital theory, 377–78
  and planning illustrations, 372–76
  problem areas in, 378–79
Capital structures of corporations
  as affected by conservatism, 420
  debt vs. equity capital, 421–22
Cobb-Douglas function, 212–13
Coefficient of multiple determination, 69
Competition; *see* Antitrust laws
Contribution profit, 112–13
Coordinative function, 4, 211, 218–19
Correlation analysis; *see also* Forecasting; Economic measurement
  and coefficient of multiple determination, 69
  deflated data in, 81
  first differences vs. actual data in, 77–79
  in forecasting, 65–69, 75–82, 117
  graphic method of, 58–70
  and intercorrelation, 81–82
  scatter diagram in, 60–63
  time lags and trends in, 80–81
Cost
  advertising or selling, 268–73
    and long-run budgeting, 271–73
    and short-run budgeting, 268–71
  of capital; *see* Cut-off rate
  classification of, 234–42
  distribution, 273–76
    types of, 275–76
  economic or opportunity, 94–96, 234, 403
  measurement of
    assessment of empirical methods of, 265–68
    methods of, 248–49
    problem areas in, 249–54
    studies in, 254–64
Cut-off rate
  alloted funds, 400–402
    debt aversion, 401
  cost of capital
    borrowing rate as, 403
    and capital structure, 407–9
    confusion regarding, 402
    defined, 409
    earnings yield as, 404–5
    lending rate as, 403
    and leveraging, 406–7

Cut-off rate—*Cont.*
  cost of capital—*Cont.*
    an opportunity cost, 403
    patterns of, 413–20
      determinants of, 415–20
    program for reduction of, 422–36
      corporate income tax, 426–27
      dividend policy, 424–25
      public relations policy, 432–36
      retained earnings policy, 425–26
    and uncertainty, 410–11

D

Debt aversion, 401
Demand
  and advertising, 154–60
  and consumption function, 147–49
  cross elasticity of, 152–53
  elasticity, 139–46
    of substitution, 153–54
  and income, 146–47, 150–51
  market testing of, 158–61
  measurement of, 135–61
    statistical consideration in, 137–39
  for substitutes, 151–54
Demand forecasting, econometric studies in
  for capital goods, 189–97
    portland cement, 194–96
    steel, 191–94
      Whitman study, 191–92
      Yntema study, 192–94
  for consumer durables, 173–89
    appliances and furniture, 180–89
      refrigerators, 182–86
      television sets, 186–88
    automobiles, 177–80
  for consumer nondurables, 161–73
    beer, 166–68
    gasoline, 162–66
    meat, 170–73
    women's dresses, 168–70
Depreciation
  defined, 97
  measurement of, 97–100
  and price-level changes, 106–9
  and tax policy, 101–3
Differential pricing, 289–91
  conditions for, 307
  degrees of, 306–7
  kinds of, 307–18
    geographic, 310–15
    product use, 317–18
    quantity, 308–10
    time, 315–17
Diminishing returns, law of, 205–10
Discounting future flows; *see* Profit, over time; Capital expenditure criteria
Dynamics, 118–24; *see also* Profit, dynamic aspects of

E

Econometrics, 39–47; *see also* Demand forecasting
Economic concentration, measurement of, 362–65
Economic measurement, methods of
  accounting and engineering, 56–57
  controlled experiments, 54–56
    Latin square, 54–55
  correlation analysis, 58
  sample surveys, 52–54
Elasticity
  of demand, 139
  of productivity, 208–9
  of substitution, 153–54

F

Forecasting, methods of
  econometrics
    correlation, 42–44, 65–70, 75–82
    statistical aspects of, 40–42
    types of models, 44–47
  factors in choice of, 47–48
  lead-lag series, 33–34
  naive, 26–33
    continuity models, 28–29
    cyclic models, 29–31
    factor listing, 26–28
    time series projections, 31–33
  opinion polling, 35–39
    Fortune poll, 36
    McGraw-Hill survey, 36
    Survey Research Center, 36–37
  pressure indexes, 34–35

G

Graphic correlation; *see* Correlation analysis

I

Incremental profit, 123–24
Independency; *see* Randomness
Indivisibility, 211
Innovation theory, 89–90, 93–94
Intercorrelation, 81–82
Interest rate, 403–4
Inventory valuation
  by FIFO, 103–4
  by LIFO, 104–6
Investment timing; *see* Reinvestment problem
Isoquants, 216–17

K

Kinked demand curve, 283–84
Kurtosis, 15

L

Latin square, 54–56
Lending rate, 403–4
Loss leaders, 303

M

Marginal efficiency of capital, 389
Marketing analysis, 156–61
Multiple products; *see* Product line
Multiple relations, 73–75, 135–36, 210–19

N

Naive method of forecasting, 26–33

O

Operations research, 226–30
  and linear programming, 227–30
Opinion polling, 35–39
Outlay-revenue ratio, 390–93

P

Patents, 332–37
Payout, 381–82
Payout reciprocal, 389–90, 393–95
Prediction; *see also* Forecasting
  a priori deduction, 5–6
  empirical measurement, 5–6
  objective and subjective, 5–8
Price discrimination, 339–52
Price leadership, 298–99
Price-level changes, effect of, 103–9
Price lining, 285–87
Pricing; *see also* Product line pricing
  "at the market," 284
  customary, 283–84
  differential, 289–91, 306–19
  methods of, 291–300
    cost-plus, 292–94
    experimental, 296
    intuitive, 295–96
    stable and initiative, 296–99
      price leadership, 298–99
  odd, 281–82
  prestige, 284–85
  psychological, 282–83
  theory of, 280–81
Probability; *see also* Uncertainty, types of
  a priori, 6
Product line
  contraction, 225–26
  and excess capacity, 220–22, 224–25
  expansion, 224–25
  interdependence, 224
  optimum, 222–23, 228
  policy, 219–26
  pricing, 301–6
    of complementary goods, 302–5
      loss leaders, 303
      tie-in sales, 304
      two-part tariff, 304–5
    of substitute goods, 301–2
Production
  functions, 202–19
  stages of, 205–9
Profit
  accounting vs. economic, 94–96
  contribution, 112–13
  distribution of, 91–93
  dynamic aspects of, 118–24
  forecasting and control of, 109–17
  incremental, 123–24
  limiting factors, 125–27
  marginal, 122–23
  maximization of, 120–24, 218–19
  measurement of, 94–108
  normal, 94–96
  over time, 118–20
  planned and unplanned, 117–18
  policies, 124–32
  standards, 127–32
  theories of, 87–94
  true rates of, 94, 380–89
  under static conditions, 120–21

Q

Quantity discounts, 288

R

Randomness, 6, 41
Rate-of-return analysis; *see also* Capital expenditure criteria
  and reinvestment problem, 427–32
Regression line, 62–65
Resale price maintenance, 287–88, 355–59
Restrictive agreements, 329–30
Returns to scale
  constant, 210–12
  law of, 210
Risk
  decision implications of, 7
  defined, 5
  insurability of, 5–8, 22
  interfirm, 8
  intrafirm, 7
  objective, 5–7
  and objective prediction, 5–8
  subjective, 10

S

Sample surveys, 52–54
Scatter diagram, 60–63
Sequential decisions, 16–18, 53–54
Simple relations, 71–72, 135–36, 202–10
Skewness, 14
Statics, 117–21

T

Tariff, two-part, 304–5
Tie-in sales, 304
Time series
  adjustments for, 137–38
  projections of, 31–33, 116
Trade marks, 337–38
Tying contracts, 338–39

U

Uncertainty; *see also* Profit, dynamic aspects of
  and antitrust enforcement, 329–61
  in capital planning, 379, 411–12
  decision making under, 4–24
    areas of, 19–21
Uncertainty—*Cont.*
  decision making under—*Cont.*
    degrees of, 13–15
  and risk, 5–8
    objective prediction, 5–8
    subjective prediction, 8
  types of, 9-13

This book has been set on the Linotype in 9 and 10 point Janson, leaded 2 points. Chapter numbers and titles are in 18 point Spartan Medium. The size of the type page is 27 by 47 picas.